

# **B** I O M E T R I C S

*March 1948*

**Vol. 4 NO. 1**

**THE BIOMETRICS SECTION, AMERICAN STATISTICAL ASSOCIATION**

---

**A Quantitative Theory of Genetic  
Recombination and Chiasma Formation**

*R. A. Fisher*

**La Relation d'Allométrie:  
sa Signification Statistique et Biologique**

*Georges Teissier*

**The General Theory of Prime Power Lattice Designs**

**I. Introduction and Designs for  $p^n$  Varieties in  
Block of  $p$  Plots**

*Oscar Kempthorne and  
Walter T. Federer*

---

---

Material for *Biometrics* should be addressed to the Chairman of the Editorial Committee, Institute of Statistics, North Carolina State College, Raleigh, N. C.; and material for Queries should go to "Queries", Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.

---

#### Officers

##### American Statistical Association

President, George W. Snedecor; Directors, Isador Lubin, Lowell J. Reed, Walter A. Shewhart, Samuel A. Stouffer, Frederick F. Stephan, Willard L. Thorp; Vice-Presidents, A. Ford Hinrichs, Rensis Likert, Abraham Wald, Harry R. Wellman; Secretary-Treasurer, Lester S. Kellogg; Acting Secretary-Treasurer, Sylvia Castleton Weyl.

#### Officers

##### Biometrics Section

Chairman, Joseph Berkson; Secretary, John H. Watkins; Section Committee, D. J. Finney, Margaret Merrell, P. J. Rulon and J. W. Tukey.

#### Editorial Committee

##### *Biometrics*

Chairman, Gertrude M. Cox; Members, C. I. Bliss, W. G. Cochran, Churchill Eisenhart, J. W. Fertig, H. C. Fryer, Horace Norton, A. M. Mood, G. W. Snedecor, and Jane Worcester.

---

Membership dues in the American Statistical Association are \$8.00 a year, of which \$4.00 is for a year's subscription to the quarterly *Journal*, \$.75 is for a year's subscription to *The American Statistician*. Members of The Association may subscribe to *Biometrics* at a special rate of \$2.00. Dues for associate members of the Biometrics Section are \$3.00 a year, of which \$2.00 is for a year's subscription to *Biometrics*. Single copies of *Biometrics* are \$1.00 each and annual subscriptions are \$3.00. Subscriptions and applications for membership should be sent to the American Statistical Association, 1003 K Street, N. W., Washington 6, D. C.

Entered as second-class matter, May 25, 1945, at the post office at Washington, D. C., under the Act of March 3, 1879. *Biometrics* is published four times a year—in March, June, September and December—by the American Statistical Association for its Biometrics Section. Editorial Office: 1603 K Street, N. W., Washington 6, D. C.

---

# B IOMETRICS The Biometrics Section of the American Statistical Association

---

Number 1	March 1948	Volume 4
----------	------------	----------

---

A Quantitative Theory of Genetic Recombination and Chiasma Formation . . . . .	<i>R. A. Fisher</i>	1
Discussion of Fisher's Paper . . . . .	<i>D. G. Catcheside</i>	9
	<i>Hilda Geiringer</i>	10
	<i>Alexander Weinstein</i>	11
	<i>J. Lederberg</i>	12
	<i>R. A. Fisher</i>	13
La Relation d'Allométrie : sa Signification Statistique et Biologique	<i>Georges Teissier</i>	14
Discussion of Teissier's Paper . . . . .	<i>Lois M. Zucker</i>	48
	<i>Harold Hotelling</i>	50
	<i>N. Rashevsky</i>	51
	<i>Georges Teissier</i>	52
The General Theory of Prime-Power Lattice Designs		
I. Introduction and Designs for $p^n$ Varieties in Blocks of $p$ Plots . . . . .	<i>Oscar Kempthorne and Walter T. Federer</i>	54
Queries . . . . .		80
The Biometric Society . . . . .		82
Annual Meeting of the Biometrics Section . . . . .		86
Statistical Summer Sessions at the University of California . . . . .		87
News and Notes . . . . .		89



# A QUANTITATIVE THEORY OF GENETIC RECOMBINATION AND CHIASMA FORMATION\*

R. A. FISHER

*Cambridge University*

## 1. INCOMPLETENESS OF GENETICAL THEORY

I have undertaken to talk on the quantitative theory of the formation of chiasmata and of genetic recombination with some diffidence, because the ideas I shall try to put before you are so new as to be in some respects immature, or half developed. They have arisen in connection with the experimental work with mice carried out in the Department of Genetics at Cambridge during the last few years. I can reasonably attempt in the course of one lecture to give the background of reasoning on which these ideas are based, but there is one difficulty which I cannot hope to overcome, and that is to give the detailed mathematical developments, largely due to my colleague George Owen, by which the primary scheme is connected with its detailed and experimentally verifiable consequences.

The subject of crossing over has for many years seemed to me to be that part of genetics in which teachers are furthest from having a complete theory to present. Let me illustrate this as simply as I can. We may think of a strand or chromosome in a gamete, having on it two points or loci marked by genetic factors. Then, in the preceding meiosis the number of breaks in the intervening tract, involving interchange of parts of the parental chromosomes, is one or another of the numbers:—

0, 1, 2, 3, . . . .

These events are mutually exclusive, and if their respective

---

\* Lecture delivered September 5, 1947, before the First International Biometric Conference.

probabilities are represented by:

$$p_0, p_1, p_2, p_3, \dots ;$$

then, since one or another of these events must have occurred, it follows that the sum of their probabilities is unity, or that

$$p_0 + p_1 + p_2 + p_3 + \dots = 1.$$

That is simple and obvious, but I need not remind geneticists that this simple specification of the rules of gamete formation is not the usual way in which the subject has been discussed. It requires no mention to be made of the possible configurations of the four-strand stage. It specifies, however, all that is relevant to the purely genetic question:—what kinds of offspring will be produced by any given kind of parent. The genetically observable, or quasi-observable, quantities can in fact be specified in terms of the series of  $p$  values.

First, the recombination fraction is directly observable by breeding tests; it is evidently equal to the sum of the probabilities of an odd number of breaks. I.e.—

$$y = p_1 + p_3 + p_5 + \dots$$

Next, the map distance,  $x$ , may be defined as the expected number of breaks in the tract, so that

$$x = p_1 + 2p_2 + 3p_3 + 4p_4 + \dots$$

These relations may be expressed somewhat more clearly by defining a second series of quantities  $s_r$  to represent the frequencies of  $r$  or more interchanges in the segment, so that

$$s_r = \sum_{i=r}^{\infty} p_i, \quad p_r = s_r - s_{r+1}$$

Then, the equations,

$$y = s_1 - s_2 + s_3 - s_4 + \dots$$

$$x = s_1 + s_2 + s_3 + s_4 + \dots$$

express the two genetic observables in terms of the  $s$ -series.

Now, any adequate theory of crossing over must supply the means of expressing the entire series of  $p$ -values, or of  $s$ -values, in terms of the positions of the genetic markers on the chromosome, and their relationships to its centromere and to its ends. How far genetical theory is in fact from meeting this requirement may be illustrated by the attempts,

of obvious practical value, to express  $x$  in terms of  $y$ . Since only  $y$  is directly observable this is a primary need in making a linkage map. Early attempts at this problem were made by Haldane, and a few years ago a remarkably effectual formula was put forward by the Indian mathematician, Kosambi, namely that

$$2y = \tanh(2x).$$

Kosambi's formula leads to an addition theorem for the recombination fraction of the sum of 2 segments,

$$y_{1+2} = \frac{y_1 + y_2}{1 + 4y_1y_2},$$

and with three genetic markers this may be tested by direct comparison with the data. In such tests Kosambi's formula usually shows up very well, yet it can scarcely be accurate in all cases, for it takes no account of the position of the segment in relation to the centromere, a factor which has long been known greatly to influence interference.

In the second place, even if it were accurate in expressing a fixed relationship between  $x$  and  $y$ , a formula of this kind could not supply a complete theory. For work with three factors the relationship between  $x$  and  $y$  is sufficient, but with four or more factors segregating simultaneously it is easily seen to be insufficient. With four factors, for example, there are eight pairs of complementary genotypes, and seven constants are needed to specify their relative frequencies; yet an addition formula, such as that given above, can only give the recombination fractions between the six pairs of loci. A further parameter is needed to specify the gametic series. In fact, only the complete series of expressions for  $s_1, s_2, s_3 \dots$  in terms of some common parameters can give a theoretically complete account of recombination.

## 2. MATHER'S SKETCH OF CROSSING OVER AS A RANDOM PROCESS

I should like to recall at this point some quite early proposals made by K. Mather in which the formation of chiasmata in a four strand pair of chromosomes was regarded as a random process, starting from the centromere, in such a way that the interval between the centromere and the chiasma next to it was assigned a definite frequency distribution, and the interval between the first chiasma and the second was treated in the same way. It is evident that with proper specification of these distributions, calculation should be able to provide the frequencies of all definable configurations, specifying the incidence of

simultaneous sets of chiasmata in the same chromosome.

I have followed Mather's ideas in this matter with some variations. First, I consider only the configurations provided by sets of chiasmata on the same strand, for these are enough to specify the genetic situation. Next, it appears that good agreement with the facts may be obtained by introducing a metric which is not to be identified with map distance, or with physical length, in terms of which interference is a constant property at all parts of the strand. In this metric it appears to be sufficient, at least as a good approximation, to treat the centromere as if it were an obligatory chiasma. In terms of this metric it is then theoretically possible to calculate the series  $s_1, s_2, s_3, \dots$  for any chosen points, and in particular the values of  $x$  and  $y$ , so as to give full information about recombination for all regions of a given map.

### 3. THE SEX CHROMOSOME OF THE HOUSE MOUSE

Some years ago, when Mather was working with me at the Galton laboratory, we carried out, with the able assistance of Mrs. S. Holt, an experiment which at the time must have seemed rather extravagant. Although the laboratory rodents had long been the best studied species among the mammals, yet sex linkage had not appeared among them. In these circumstances, it seemed to us worth trying whether any of the factors we had readily available were, as was possible, carried in the pairing segment of the sex chromosome, and in consequence exhibited partial sex linkage. In the event, none of the six factors we tested: pied, agouti, brown, dilute, wavy ( $wv_1$ ), and light head, appeared to be linked with sex in a test with about twelve hundred mice. Yet before our experiment was ready for publication, J. B. S. Haldane, working in the same College had made a search for partial sex linkage in Man and had found pedigrees of several rare defects appearing to be so inherited. The test for a similar phenomenon in mice, was evidently not so hopeless.

When in 1943 I moved to Cambridge I was able, through the generosity of The Rockefeller Foundation and the kindness of Dr. Snell of the Roscoe B. Jackson laboratories to increase the variety of my mouse stocks, and, about two years ago Mrs. M. Wallace, then Miss Margaret Wright, noticed a certain irregularity in the sex ratios of some of the lines which we kept segregating for the linked factors wavy-two and shaker-two. In consequence we decided to set up an experiment which we hoped would be competent to discover what was really happening

at these two loci. The effect to be demonstrated was evidently not very large, and there were reasons to fear, as there usually are with recessive factors, and especially shakers, that the evidence would be disturbed by differential viability. Consequently, the experiment had to be carefully designed.

There are four possible types of male heterozygous in these 2 factors, for the animal may have received from his father, either wavy or shaker or both or neither, with a complementary set of genes from the mother. Of these four male types that in which neither recessive comes from the father and both from the mother, i.e.

$$wv_2 sh_2 X / ++ Y \quad (\text{Tricoupling})$$

may be regarded as in coupling for all three factor pairs, while the other three types differ from it respectively by interchanging sex ( $X$ ,  $Y$ ), wavy ( $+$ ,  $wv_2$ ), and shaker ( $+$ ,  $sh_2$ ) giving therefore

$$++ X / wv_2 sh_2 Y$$

$$+ sh_2 X / wv_2 + Y$$

$$wv_2 + X / + sh_2 Y$$

Each of these male types mated with homozygous wavy shaker does,

$$wv_2 sh_2 X / wv_2 sh_2 X$$

produces eight genotypes, four male and four female, which fall into four pairs of complementary genotypes

Wavy shaker females and normal males	A
Normal females and wavy shaker males	B
Shaker females and wavy males	C
Wavy females and shaker males	D

The pairs of genotypes B, C, D are derived from the pair A, by the same three interchanges of male with female, of normal with wavy, and of normal with shaker.

In relation to any one of the types of heterozygous fathers these four pairs of genotypes are produced by four *modes of gamete formation* of which one, in the case of three loci in the same chromosome, may be identified as old combinations, while in the others recombination has taken place in one segment or the other, or in both. Without knowing the order of the loci we cannot tell which of these is which; but in all cases the three modes of formation involving recombination will differ

from the first, in which no recombination has occurred, by the three operative interchanges for sex, wavy and shaker.

Consequently the data obtained when all four male types are used will consist of sixteen frequencies, each of a pair of complementary genotypes arranged in a symmetrical four by four Latin square, as shown below:

Type of gamete formation

Type of male	No change	Sex change	Wavy change	Shaker change
$wv_2 sh_2 X/++Y$	A	B	C	D
$++X/wv_2 sh_2 Y$	B	A	D	C
$+sh_2 X/wv_2 Y$	C	D	A	B
$wv_2 X/+sh_2 Y$	D	C	B	A

The experimental purpose of obtaining data in this form is to allow of the estimation of the frequencies of the four modes of gamete formation, free from disturbances introduced by unknown differences in the viabilities of the four pairs of genotypes, and of arbitrary differences in the numbers of offspring bred to the four types of male. This statistical problem is intricate and has never been adequately discussed. Fortunately, however, the particular data obtained with the factors wavy-two and shaker-two in mice offer no difficulty, for there were in our results no signs of differential viability, as appeared from the complete homogeneity of the four distributions shown by the rows of the table.

The relative frequencies of the four modes of gamete formation may in such a case be estimated directly from the total frequencies in the columns. Out of about 450 mice bred it was found that the first column contained about 29%, the second 41%, and the remaining columns about 15% each. I need not emphasize that such a result appeared highly paradoxical. For, in the first place, if sex linkage were not involved, we should expect equality in the first and second columns, as we would also in the third and fourth; whereas the first two columns differ consistently for all types of male, and in the aggregate by over 50 mice. Next, if sex-linkage were involved we should expect entries either in the third or in the fourth column to be very rare double cross-overs, but these two classes are nearly equally frequent. Finally with incomplete sex-linkage we should expect the sex interchanges to be less numerous than the cases of no recombination.

In respect to recombination frequencies we have:—

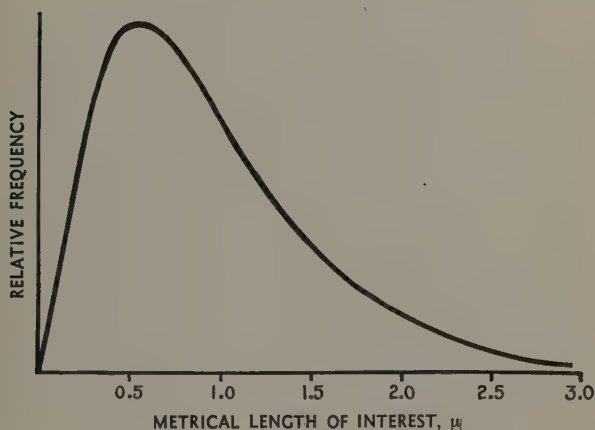
		More accurately
Wavy-shaker	$15 + 15 = 30\%$	$31.1\% \pm 2.17$
Sex-shaker	$41 + 15 = 56\%$	$56.1\% \pm 2.33$
Sex-wavy	$41 + 15 = 56\%$	$56.7\% \pm 2.33$

So that the paradoxical nature of our data may be restated by saying that both the locus for wavy, and that for shaker, have significantly more than 50% recombination with sex; moreover, these recombination fractions are nearly equal, although the loci must be 30 to 35 map units of distance apart.

#### 4. TOPOGRAPHICAL INTERPRETATION OF THE DATA

In the male mouse it is to be supposed that the inheritance of sex takes place as if a sex determining locus were situated at the point of separation of the pairing segment from the differential segment. For

Figure 1  
Frequency Distribution of Length of  
Intercept Between Successive Breaks



purposes of calculation it is convenient to take this point as origin. The occurrence over a considerable length of the pairing segment of recombinations with sex exceeding 50 percent, becomes more intelligible if we consider that the complete absence of chiasmata in the differential segment of these strands must mean an absence of interference from this side; so that, on a metric chosen for uniformity in respect of inter-

ference, this region will be favored by chiasmata. Moreover, calculation taking any plausible form (see fig. 1) for the distribution of the intercept length such as assigning to it the frequency element

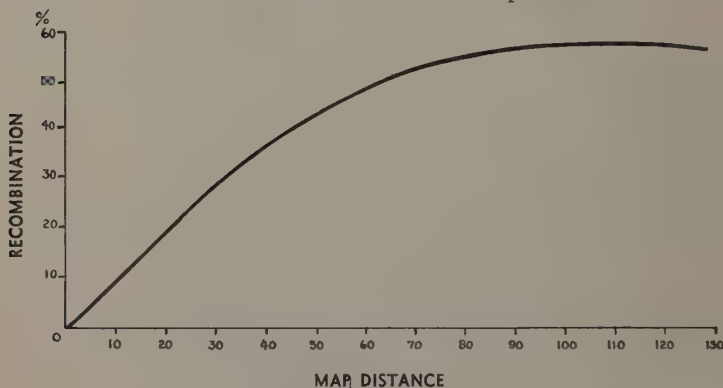
$$\operatorname{sech} \left( \frac{1}{2} \pi u \right) \tanh \left( \frac{1}{2} \pi u \right) d \left( \frac{1}{2} \pi d \right)$$

where  $u$  is the metrical value, and the coefficient  $\frac{1}{2} \pi$  is introduced to

make the metric conform to map distance in regions far from the arm ends, or from the centromere, shows that this expected effect is realized.

If now numerical values of  $s_1, s_2, s_3, \dots$  are tabulated for various distances from the end, the value of the sex recombination fraction,  $y$ , and the map distance,  $x$ , may each be obtained in terms of  $u$ . Then, plotting the recombination fraction  $y$  against the map distance  $x$ , it appears that  $y$  rises at first in equality with  $x$ , but passes the 50 percent value at a map distance of about 60 map units and rises to a long flat maximum between 55% and 56% recombination.

Figure 2  
Recombination Fraction in Terms of Map Distance



The genetical findings with wavy and shaker would therefore be predicted by theory if these two loci were situated about 30 units apart in the neighborhood of the maximum recombination fraction and probably on either side of it.

In a more detailed study George Owen has shown how the effects of finite lengths of the chromosome arms may be taken into account; from

his work it appears also that recombination fractions exceeding 50 per cent are characteristic of a wide class of intercept distribution functions, provided interference is sufficiently pronounced. He has found models in which explicit analytic functions may be used to replace the somewhat cumbrous numerical integration I had employed.

As I mentioned in my opening remarks, very much detailed work remains to be done, although, so far as we have explored the question, the genetical data have shown themselves entirely conformable with the numerical details of the theory. The exact degree of confirmation obtainable from the available data for *Drosophila* will, however, not be clear until a more thorough study has been made of the relationship of coincidence values, appropriately estimated, to the chromosome regions in which they are found.

## DISCUSSION

**D. G. Catcheside.** The occurrence of recombination in excess of 50 per cent requires that there should be chromatid interference favoring a frequency of four-strand doubles in excess of that expected in the absence of chromatid interference.

When two successive chiasmata occur in an arm of a pair of chromosomes, the relationships between the four chromatids may be such that two strands (complementary), or three strands (diagonal) or four strands (reciprocal) take part in the crossing-over at the two chiasmata. If the participation of a given strand in one chiasma does not affect the chance of its participation in the other chiasma in the arm, the strand relationships should occur in double crossing-over with the frequencies of a quarter complementary, a half diagonal and a quarter reciprocal. This absence of chromatid interference is found in *Drosophila melanogaster* and with such absence the amount of recombination between linked genes cannot rise about 50 per cent.

Chromatid interference may be of two types; either (a) the participation of a chromatid in one chiasma reduces its chance of participation in a second chiasma (positive interference) or (b) participation of a chromatid in one chiasma increases its chance of participation in a second chiasma (negative interference). The latter, in which the upper limit of recombination would be less than 50 per cent, has been

observed in *Neurospora crassa*, where 2-strand doubles are in excess and 4-strand doubles are deficient. The former type, which could in some circumstances lead to an amount of recombination between two loci in excess of 50 per cent, has not so far been observed. It may be that crossing-over in the heteromorphic pair of sex chromosomes in the male mouse is of this type, but it must not be assumed that positive chromatid interference is of general occurrence.

While it is true that most genetical experiments have shown an absence of chiasma interference across the centromere, recent work of Callan and Montalenti (*Journal of Genetics* 48, 119; 1947) has confirmed the occurrence in *Culex pipiens* of positive chiasma interference across the centromere as previously recorded by Pâtan. Thus the centromere cannot always be taken as a null point.

**Hilda Geiringer.** Professor Fisher's new and extremely interesting theory of recombination suggests the following question: The complete linkage of  $m$  factors is characterized by  $M = 2^{m-1} - 1$  parameters. In fact, denote the constitution of an organism by  $(x_1, x_2, \dots, x_m; y_1, y_2, \dots, y_m)$  where the maternal (paternal) heritage with respect to  $i$ -th factor is denoted by  $x_i$  or  $y_i$  respectively. A gamete may then be described, for example, by  $x_1, x_2, y_3, y_4, y_5, \dots, x_m$ ; there are  $2^m$  gametes of this sort and there exists a probability distribution corresponding to these  $2^m$  possibilities. Admitting obvious symmetry relations this number,  $2^m$ , will be divided by two and we thus have  $2^{m-1}$  probability values with sum equal to one, hence  $M$  parameters. This "linkage distribution" (l.d.) is completely equivalent to the cross-over distribution, or "recombination distribution" (r.d.) which (in the "linear theory") specifies the probabilities that in each of the  $(m-1)$  intervals between the loci of linked factors recombination takes place or does not take place.

On the other hand this r.d. is, in general, not equivalent to the  $N = m(m-1)/2$  recombination values. In fact  $M > N$  for  $m \geq 4$ . From the r.d. (or from the l.d.) the recombination values follow, but not vice versa. It is, however, an aim to be able (by means of observation and theory) to reduce the number  $M$  of independent parameters and to express the  $M$  values by means of the  $N$  recombination values or possibly by part of them, e.g. by the  $(m-1)$  recombination values between adjacent factors. May I ask Professor Fisher's opinion regarding this question, in the light of his new theory?

**Alexander Weinstein.** *Recombination exceeding fifty per cent.* The recombination frequency between two loci depends on (1) the number of chromatids that cross over at a level, (2) which chromatids these are (whether homologous or sister strands), (3) the number of levels at which crossing over occurs, and (4) the relation of exchanges at different levels to one another (whether regressive, progressive, or digressive).

If at any level only two chromatids of a tetrad cross over and these are always homologous and not sister strands, then, on certain assumptions, in crossover tetrads with  $n$  exchanges between two loci, the proportion of chromatids showing recombination between these loci is  $\frac{1}{2}$  when  $n$  is odd and  $\frac{1}{2} - \frac{1}{2} (F - H)^{n/2}$  when  $n$  is even; where  $F$  is the chance that an exchange is regressive and  $H$  the chance that it is digressive with respect to an adjacent exchange.<sup>1</sup> These formulas apply not to chromatids from all tetrads taken together, but to chromatids from tetrads of a specified rank  $n$ .

If  $F = H$ , then in a group of crossover tetrads of rank  $n$ , the recombination frequency of emerging chromatids remains  $\frac{1}{2}$  as  $n$  increases. If  $F$  exceeds  $H$ , the frequency will oscillate between  $\frac{1}{2}$  and a lesser value which approaches  $\frac{1}{2}$  as  $n$  increases indefinitely. If  $F$  is less than  $H$ , the recombination frequency oscillates between values greater and values less than  $\frac{1}{2}$ , and these values approach  $\frac{1}{2}$  as  $n$  increases indefinitely. That is, the recombination frequency in chromatids derived from tetrads of rank  $n$  exceeds  $\frac{1}{2}$  if  $F$  is less than  $H$  and  $\frac{n}{2}$  is odd; and if such tetrads are sufficiently numerous, the recombination frequency in chromatids from all tetrads taken together may exceed 50 per cent.

These results assume that  $F$  and  $H$  are constant throughout the chromosome. If either or both vary according to the regions involved, the recombination frequency in chromatids derived from tetrads with exchanges in  $n$  specified regions between two loci remains  $\frac{1}{2}$  when  $n$  is odd; and when  $n$  is even, the recombination frequency becomes  $\frac{1}{2} - \frac{1}{2} (F_2 - H_2) (F_4 - H_4) (F_6 - H_6) \dots (F_n - H_n)$ , where the subscript indicates the second of the two adjacent regions involved; for example,  $F_4$  is the chance that an exchange in region 4 is regressive with respect to an exchange in region 3. The recombination frequency in chromatids derived from a crossover tetrad of rank  $n$  would exceed  $\frac{1}{2}$  if  $F$

---

<sup>1</sup> Weinstein, Alexander, "Mathematical study of multiple-strand crossing over and coincidence in the chromosomes of *Drosophila*," *American Philosophical Society Year Book* 1937, pp. 227-228, 1938.

is less than  $H$  in an odd number of factors, provided that in the other factors  $F$  exceeds  $H$ . If  $F$  equals  $H$  in one or more factors, the recombination frequency is  $\frac{1}{2}$ .

The recombination frequency might also exceed 50 per cent if an exchange between two chromatids of a tetrad were accompanied by an exchange between the two others at the same level. This would yield 100 per cent recombination in chromatids derived from tetrads with crossing over at one level: the result would be the same as if the two exchanges were at different levels and digressive with respect to each other. If two exchanges at the same level are treated as mutually digressive, the formulas already given can be applied, whatever the number of levels at which there is crossing over; provided that  $n$  is taken as the number of exchanges, which is now no longer identical with the number of levels at which exchanges occur.

The relative frequency of regressive, progressive, and digressive crossing over also affects other genetic functions, including coincidence of various types. In attached  $X$ -chromosomes, the frequency of homozygosis in individuals derived from tetrads of rank 2 or higher depends on the frequency of progressive exchanges,  $G$ , which equals  $1 - F - H$ . Thus in suitable material the values of  $F$  and  $H$  calculated from the recombination frequencies can be subjected to additional tests.

(Such a notation was presented orally, but is omitted from the written proceedings subject to publication elsewhere.)

**J. Lederberg.** Professor Fisher has given us a strong justification of what he terms a genetic as opposed to a cytogenetic approach to the problem of interference in crossing-over. We shall, however, all be interested to see in the further development of this analysis the extent to which his genetic formulation is consistent with, or even relevant to, cytogenetic data on interference in the four-stranded bivalent in which the physical phenomenon of crossing over is believed to occur.

In bridging this gap, the usual procedure is the construction of many tetrad diagrams in which the effects on the gametic output of combinations of crossovers involving various strands at different levels are worked out geometrically. The geometry is, however, sufficiently simple to be amenable to convenient notation, by the use of which the effects of crossing over can be elucidated algebraically. An algebra is most useful perhaps in respect to the formulation of theories in terms of restraints on the values of its terms, which may be more easily manipulated than restraints geometrically expressed.

**R. A. Fisher.** In reply to Mr. Catcheside, I am surprised and somewhat taken aback that he should be of the opinion that chromatid interference is absent in *Drosophila melanogaster*. Perhaps he only means that the interference observed could conceivably be explained without specifically chromatid interference; but this seems to be quite another matter. It is at least obvious that the *Drosophila* data accords excellently with the theory here under discussion.

In reply to Dr. GEIRINGER: The new theory does supply formulae, in some cases evaluable easily, and in others with more difficulty, which give the frequencies of all possible recombinations to be found in gametes from a multiple heterozygote. These are given in terms of the metric-values of the different segregating loci relative to the centromere, of the arm lengths of the two parts of the chromosome, and, possibly, of an additional parameter available for specifying the intensity of interference.

In respect to the points raised by Dr. WEINSTEIN and others in the discussion, I should have made it clear that my theory does not depend on the whole complexity of the configurations possible in the four-strand stage, but only on the frequency distribution of the different configurations possible to a single strand received by a gamete. This simpler distribution is sufficient for the purely genetic purpose of predicting the frequencies of the different genotypes derived from a multiple heterozygous parent.

# LA RELATION D'ALLOMÉTRIE SA SIGNIFICATION STATISTIQUE ET BIOLOGIQUE\*

GEORGES TEISSIER

*Station Biologique de Roscoff*

Le chapitre de l'étude de la croissance dont les progrès récents ont été les plus remarquables est sans contredit celui qui traite des changements de proportions des différents organes ou, autrement dit, de la croissance relative. Les recherches sur cette question, fort nombreuses depuis une vingtaine d'années, ont fourni matière à plusieurs mises au point, inégalement développées.<sup>1</sup>

La question est actuellement si vaste et si complexe, qu'il apparaît qu'en un exposé limité, il faut se contenter, ou de ne traiter que les grandes lignes du sujet, en laissant de côté, non seulement bien des points importants, mais aussi la plupart des difficultés qui y figurent encore, ou de s'attacher à une question beaucoup plus limitée, en essayant de l'approfondir. C'est à ce dernier parti que je me suis arrêté et il ne sera question ci-après que d'un problème unique, qui, à la vérité, se trouve au centre même du sujet. Dans ce qui va suivre, nous allons chercher la signification exacte de la *relation d'allométrie* qui, depuis plus de vingt ans est l'instrument indispensable de toute étude sur la croissance relative.

Le fait fondamental en la matière est l'existence d'un type de relation qui, convenablement ajusté, permet de décrire exactement les croissances relatives les plus diverses. Chez tous les animaux appartenant à tous les grands embranchements étudiés jusqu'à présent, les relations entre, d'une part, les dimensions ou le poids d'organes très différents par leur nature, leur structure ou leur fonctionnement et, d'autre part, les dimensions ou le poids du corps tout entier, ou d'un organe pris comme référence, peuvent être traduites par des équations de la forme :

$$y = bx^a$$

---

\* Lecture delivered September, 1947, before the First International Biometric Conference.

<sup>1</sup> G. Teissier, *Travaux Station Biologique de Roscoff*, 1931, 9; J. S. Huxley, *Problems of Relative Growth*, London, 1932; G. Teissier, *Act. Sc. Hermann* 95, 1934; *Annales Physiologie et Physico-Chimie Biologique*, 12, 1936, et *Act. Sc. Hermann*, 1936; E. C. R. Reeve et J. S. Huxley; P. B. Medawar; O. W. Richards et A. J. Kavanagh, *Essays on Growth and Form*, Oxford, 1945.

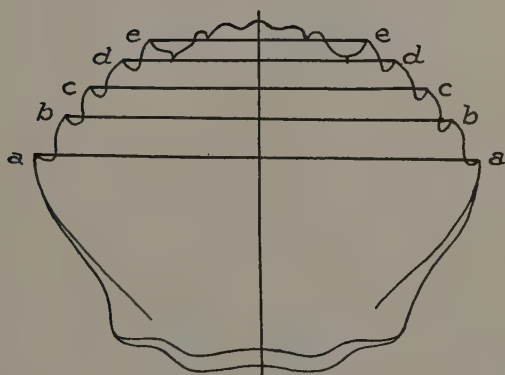
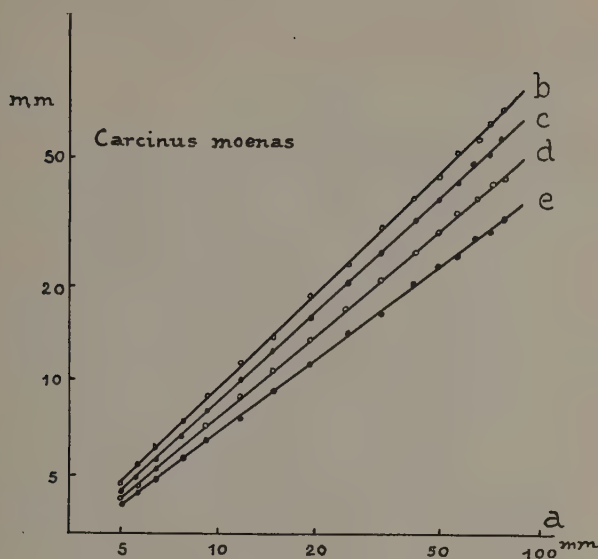


Fig. 1

Allométrie de croissance du céphalothorax de *Carcinus moenas*. Le croquis donne la définition des dimensions dont les mesures sont reportées sur le graphique. Coordonnées logarithmiques.  $\alpha = 1,00$  pour *a a*;  $\alpha = 0,98$  pour *b b*;  $\alpha = 0,93$  pour *c c*;  $\alpha = 0,86$  pour *d d*;  $\alpha = 0,74$  pour *e e*.

ou  $b$  et  $a$  sont des constantes. En coordonnées logarithmiques, ces arcs de courbe puissance se traduisent par des droites de pente  $a$  :

$$\log y = \log b + a \log x$$

Lorsque la croissance d'un organe est plus rapide ou moins rapide que celle du reste du corps, ou de l'organe de référence, et que la relation précédente est vérifiée, nous disons qu'il y a *allométrie*. Si  $a > 1$ , l'allométrie est dite positive ou majorante; si  $a < 1$  elle est négative ou minorante. Dans le cas où  $a = 1$ , la croissance de l'organe et celle du corps sont parallèles: il y a *isométrie*.

Des deux paramètres qui définissent la relation d'allométrie, le plus important  $a$ , pente de la droite figurative du phénomène en coordonnées logarithmiques, sera la *constante actuelle d'équilibre*;  $b$  qui représente la taille de l'organe  $y$  lorsque  $x = 1$  a la même signification que les indices dont font usage les anthropologistes: ce sera l'*indice origine*.

Avant de commencer un examen approfondi de cette relation, il n'est pas inutile, pour justifier le temps que nous lui consacrerons, de rappeler l'importance du rôle qu'elle joue actuellement dans l'étude de la croissance relative et dans la biométrie générale.

En premier lieu, nous devons insister sur le fait qu'elle traduit en général avec la plus grande fidélité les modalités de la croissance moyenne. On pourrait le prouver par plusieurs centaines d'exemples également probants. L'étude statistique des cas pour lesquels on possède des mesures suffisamment nombreuses permet par ailleurs de démontrer que les écarts qui se produisent de part et d'autre de cette loi moyenne peuvent être considérés comme fortuits. Elle permet en outre de comparer, avec autant de sécurité, des animaux adultes appartenant soit à la même espèce, soit à des espèces différentes. C'est à ce titre qu'elle est devenue un instrument très efficace dans l'étude des séries phylogéniques ou dans la recherche des affinités des races d'une même espèce ou des espèces d'un même genre.

Certains organes ou certaines parties du corps, dont la croissance ne peut être représentée par une seule courbe puissance, suivent successivement plusieurs relations d'allométrie. Sur les graphiques logarithmiques les droites peuvent se raccorder par un point anguleux, ou être séparées par une discontinuité plus ou moins ample, qui correspond à un stade où le poids (ou la longueur) de l'organe n'est plus fonction du poids (ou de la longueur) du corps. Une courbe de croissance relative peut ainsi présenter deux sortes d'accidents et se décompose dans les deux cas en tronçons réguliers exactement définis.

Les points anguleux et les discontinuités qui jalonnent les courbes de croissance des différents organes d'un même animal ne sont pas répartis de façon quelconque, mais apparaissent au contraire simultanément dans des organes divers. La croissance d'un animal se divise ainsi naturellement en étapes distinctes. Pendant chacune de ces étapes, les divers organes et les divers constituants biochimiques croissent, l'un par rapport à l'autre, en suivant des relations d'allométrie simples, la composition chimique, la structure et la forme variant de façon continue. Les étapes successives sont séparées l'une de l'autre par des *stades critiques* de durée assez brève où, sans que le poids ou la taille augmente sensiblement, les règles—qui présidaient jusque là à la croissance changent brusquement. Ces stades critiques se marquent sur les diverses courbes de croissance relative par des points anguleux

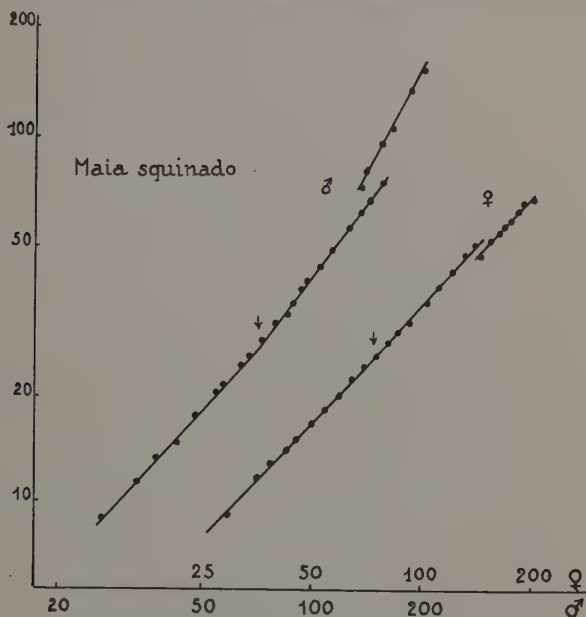


Fig. 2

Etapes de la croissance du propodite de la pince chez les *Maia squinado* mâles. Longueur du propodite en fonction de la longueur du Céphalothorax. Coordonnées logarithmiques. Mâle:  $\alpha = 1,09$  pour la première étape;  $\alpha = 1,35$  pour la deuxième;  $\alpha = 1,93$  pour la troisième. Femelle:  $\alpha = 1,09$  pour la première étape;  $\alpha = 1,16$  pour la deuxième;  $\alpha = 1,16$  pour la troisième.

ou des discontinuités. Du point de vue biochimique ils correspondent à des changements plus ou moins profonds dans l'équilibre réciproque des divers constituants de l'organisme. Du point de vue physiologique, ils correspondent fréquemment à des modifications importantes dans le jeu des corrélations humorales.

Cette conception de la croissance relative, que j'ai développée autrefois me paraît, encore aujourd'hui, correspondre assez exactement à la réalité bien que sur certains points, notamment en ce qui concerne le rôle des stades critiques, l'accord des spécialistes soit loin d'être fait. Il n'est pas question de la défendre ici mais, qu'elle soit exacte ou qu'elle soit fausse, son succès définitif, ou son échec, dépendront en définitive de la validité plus ou moins grande de la relation d'allométrie. Aucune des interprétations physiologiques, voisines d'ailleurs l'une de l'autre qui ont pu être suggérées par Huxley, par Robb, ou par moi-même, n'a jusqu'ici recueilli l'assentiment général. Peut-être en trouvera-t-on une plus satisfaisante, mais il restera encore à expliquer que la même formule convienne à tant de cas si disparates et à comprendre sa signification générale.

Avant d'aborder ce problème, une remarque préalable est indispensable. La raison essentielle du succès de la notion d'allométrie est qu'elle ne fait pas intervenir l'âge des individus dont elle compare les caractéristiques. C'est grâce à cette circonstance qu'ont pu être faites des études très précises sur la croissance d'animaux qu'il est difficile, ou même impossible, d'élever au laboratoire.

Une telle caractéristique n'est évidemment pas liée à la forme même de la relation d'allométrie et toute autre relation  $\phi(x, y) = 0$  entre deux grandeurs mesurées sur un même animal "court-circuiterait le temps" aussi bien qu'elle. Elle postule, en revanche, la réalisation préalable d'une condition rarement énoncée d'une façon explicite, parce que sans doute trop évidente, mais qui est nécessaire à la réussite d'une étude sur la croissance relative. Il faut que, chez l'organisme étudié, la forme, la structure ou la composition chimique d'un individu placé dans des conditions données, ne dépendent que de ses dimensions et non pas de son âge.

Qu'une telle condition puisse être réalisée au cours de la croissance n'est nullement une nécessité logique. Elle apparaît comme une règle pratique qui se vérifie, rigoureusement pour de nombreuses espèces, approximativement pour beaucoup d'autres, pendant la plus grande partie de leur existence et qui, chez elles, ne se trouve éventuellement en défaut qu'à certains stades critiques du développement, ceci n'ex-

cluant naturellement pas la possibilité, pour quelques espèces, de répondre par un changement de forme à certaines modifications de leur milieu. Il est encore beaucoup trop tôt pour tenter de dresser la liste des espèces appartenant à cette catégorie d'êtres à forme stable, mais on doit cependant citer au premier rang d'entre eux le grand groupe des Arthropodes. Comme par ailleurs, ces animaux se prêtent à des mesures multiples et généralement beaucoup plus précises que celles qui peuvent être pratiquées sur les représentants des autres embranchements, ils constituent, avec certains Vertébrés, un matériel de choix pour l'étude des lois de la croissance relative. Il semble, en revanche, à en juger du moins par les recherches des auteurs japonais, que les Lamellibranches se prêtent mal à ce genre d'investigation, leur forme dépendant assez largement à taille égale, non seulement des conditions de milieu, mais peut-être aussi directement de l'âge.

Nous laisserons de côté les animaux appartenant à cette deuxième catégorie pour nous occuper uniquement de ceux chez qui, pendant une partie au moins de la croissance, deux lots d'animaux de même taille et d'âge différent ne sont pas discernables l'un de l'autre et chez qui, réserve faite de la variabilité spécifique, la forme est fonction de la taille.

Il existe une autre circonstance, toute différente en son principe de la précédente, où il est légitime de comparer les dimensions  $x$  et  $y$  de deux organes en faisant abstraction du temps, c'est celle où la comparaison porte sur des animaux ayant cessé de croître et chez qui, pour un temps, ou définitivement, forme et taille sont comme figées. Il en est ainsi pour les Mammifères adultes et pour les imagos d'Insectes et l'on sait que c'est sur des animaux appartenant à ces groupes et remplissant ces conditions qu'ont été faites quelques-unes des premières applications des relations d'allométrie. Celles qui portent sur les Mammifères essaient en général de décrire les relations d'allométrie pouvant exister entre les structures d'espèces voisines mais très différentes de taille, tandis que celles qui portent sur les Insectes tentent d'analyser les différences de forme corrélatives des différences de dimensions à l'intérieur d'une même espèce. Depuis, allométrie de taille et allométrie de croissance ont été étudiées, à l'aide des mêmes méthodes, par les mêmes auteurs sans qu'aucun d'eux ait réussi à justifier entièrement l'emploi qu'il faisait d'une même formule pour traduire des relations aussi dissemblables que celles qui peuvent unir un jeune animal et l'adulte correspondant, deux adultes de la même espèce, ou deux adultes d'espèce différente. Encore moins a-t-on réussi à expliquer que la

relation d'allométrie puisse être un instrument puissant d'analyse dans l'étude des variations intraspécifiques et interspécifiques de la composition chimique des organismes.

C'est à la recherche d'une interprétation générale de la relation d'allométrie que va être consacré le présent exposé.

## I

Plusieurs schémas physiologiques, assez voisins d'ailleurs l'un de l'autre, ont été proposés pour interpréter l'allométrie de croissance. Si aucun d'eux n'a, jusqu'à présent, recueilli l'assentiment général des biologistes, il est permis d'espérer qu'un schéma analogue, mais plus parfait, sera reconnu quelque jour comme satisfaisant. Il est certain, en revanche, qu'un schéma mécaniste de ce type ne pourra s'appliquer qu'aux cas où la comparaison porte sur les stades successifs du développement d'un même individu et à ceux où les individus comparés peuvent légitimement être considérés comme fournissant l'image des différentes étapes d'une même évolution. Il perdrait toute signification si l'on voulait l'appliquer aux cas où la comparaison porte, en principe comme en fait, sur des individus différents et où la relation d'allométrie condense en une formule une infinité de possibilités dont chaque individu ne peut, dans les circonstances les plus favorables, réaliser qu'une seule. J'ai proposé en 1937, une interprétation statistique de cette *allométrie de taille* et j'ai pensé à cette époque qu'il serait déraisonnable d'essayer de la transposer au cas de l'*allométrie de croissance*. Je crois cependant aujourd'hui à la possibilité d'une interprétation des phénomènes d'allométrie qui engloberait non seulement tous les faits morphologiques, mais aussi les faits biochimiques pour lesquels il n'a, jusqu'à présent, pu être trouvé d'interprétation autre que purement formelle.

Le point de départ de cette conception unitaire est dans l'interprétation statistique de l'allométrie de taille convenablement précisée.

Soit donc une population homogène d'individus ayant atteint leur taille définitive et sur chacun desquels peut-être pratiquée avec précision la mesure d'un certain nombre de grandeurs  $x, y, z, \dots$ . Portons notre attention sur deux de ces variants  $x$  et  $y$ . Le résultat des mesures est un nuage de points  $(x, y)$  dont la forme est d'autant mieux définie que le nombre d'individus mesurés est plus grand. Nous nous limiterons aux cas, heureusement nombreux dans la pratique, où la corrélation entre les deux grandeurs considérées est forte; ce sont

d'ailleurs les seuls pour qui la notion d'allométrie puisse présenter un véritable intérêt. Le nuage figuratif de la population des mesures est alors très allongé. Il s'agit maintenant de préciser sa forme. Sur ce point, d'ailleurs capital, il n'est malheureusement pas possible d'arriver à une certitude absolue. Il se trouve que, dans bien des cas, le nuage peut être considéré comme sensiblement normal, en ce sens que ses points se répartissent à peu près suivant une loi de Laplace-Gauss à deux variables et qu'il est allongé suivant une droite. Mais il se trouve aussi que, bien souvent, les mêmes mesures traduites dans le système  $\log x$ ,  $\log y$  donnent encore un nuage de points sensiblement normal, allongé suivant une droite. A titre d'exemple de ce fait très général, je citerai les *Maia squinado* mâles adultes, chez qui le coefficient de corrélation entre la longueur du propodite de la pince et la longueur du céphalothorax est 0.957, et le coefficient de corrélation entre les logarithmes de ces grandeurs 0.965, le nuage logarithmique étant peut-être un peu mieux allongé suivant une droite que le nuage arithmétique et pouvant être, au même titre que ce dernier, considéré comme normal. L'ambiguïté de ce résultat ne tient évidemment qu'à l'insuffisance du nombre des individus mesurés et, en augmentant l'importance de l'échantillon de la population étudiée, on doit pouvoir faire le choix qui s'avère impossible sur un lot de 300 animaux, à moins que, ce faisant, on ne tombe sur un résultat incompatible avec l'une et l'autre des deux hypothèses envisagées. Dans la pratique, avec les populations de mesures dont on dispose, l'incertitude est très fréquente.

Il peut certainement arriver que le nuage de points soit plus allongé et plus rectiligne en coordonnées arithmétiques qu'en coordonnées logarithmiques et que la loi de distribution soit nettement du type Laplace-Gauss. Il est certain, en revanche, que, dans beaucoup de cas, notamment lorsqu'une des mesures porte sur un poids et l'autre sur une longueur, le nuage est plus normal lorsqu'il est construit à partir du logarithme des mesures que lorsqu'il est fait à partir de ces mesures mêmes, la distribution étant alors du type Galton-MacAlister. Mais il arrive aussi, et je l'ai constaté notamment dans l'étude des relations entre le poids des élytres et celui des mandibules chez *Lucanus cervus*, que le nuage ne soit normal, ni dans un système, ni dans l'autre, et ne s'allonge suivant une droite que dans sa partie médiane.

Laissant de côté ces distributions complexes, encore très insuffisamment connues, il apparait que la situation est la suivante : dans le plus grand nombre des cas, parmi ceux dont l'étude a été la plus complète et la plus rigoureuse, deux lois semblent pouvoir actuellement être

employées aussi légitimement l'une que l'autre. Pour pouvoir justifier, par des critères statistiques rigoureux, les choix que l'on est amené à faire entre elles, il faudrait accroître démesurément l'importance du matériel étudié. Encore n'est-il pas certain que ce faisant, on ne découvrirait pas qu'une troisième loi, plus complexe que les deux autres et intermédiaire en quelque sorte entre elles, ne conviendrait pas mieux aux faits observés. Il apparaît, dans ces conditions, que l'on est actuellement en droit d'exprimer la relation stochastique entre les deux variables par une régression logarithmique aussi bien que par une régression linéaire, constatation qui légitime en une certaine mesure, l'emploi de la relation d'allométrie sans d'ailleurs justifier la place qu'a prise cette relation dans la biométrie au détriment de la relation linéaire, qui peut cependant prétendre, aussi bien qu'elle, à représenter les faits observés. Bien que le fait que, dans un certain nombre de cas bien établis, la régression logarithmique donne un meilleur ajustement que la régression linéaire plaide en faveur de la relation d'allométrie, il reste à justifier par des arguments biologiques cette préférence générale. Nous les trouverons en approfondissant la notion de variabilité, telle qu'elle est donnée par l'observation de la nature d'une part, par la description statistique des faits d'autre part.

Nous rappellerons tout d'abord qu'aux yeux du biologiste la variation relative est plus significative que la variation absolue. Deux organes de taille différente d'un même animal sont considérés comme également variables, non pas si leurs dimensions présentent des fluctuations égales, mais si ces fluctuations ont la même importance relative; la variabilité s'évalue traditionnellement non pas en unités métriques mais en pourcentages. Il apparaît naturel, dans ces conditions, de dire que deux organes sont également variables si leurs coefficients de variabilité sont égaux et de mesurer par ce paramètre l'importance de cette variabilité. Il s'agit là d'une définition dont il est naturellement impossible de démontrer le bien fondé mais qui apparaît en parfait accord avec l'idée que les biologistes se sont fait de la variabilité. Elle admet que, dans des organes également variables et de dimensions moyennes différentes, l'écart type est proportionnel à la taille, fait en parfait accord avec cette constatation empirique que les coefficients de variabilité d'organes homologues d'animaux de taille très différente ne présentent pas de variations systématiques avec cette dernière et que les objets que le biologiste est porté à considérer comme également variables ont toujours sensiblement même coefficient de variabilité.

Cette définition de la variabilité a plusieurs conséquences inté-

essantes. La première apparaît dès que l'on examine les raisons qui peuvent être données pour expliquer que les variations quantitatives d'un être vivant puissent être traduites par une loi normale ou sensiblement normale. Ces raisons sont calquées sur celles qui sont utilisées pour expliquer la validité de la même loi dans d'autres domaines et plus particulièrement dans la théorie des erreurs, et il n'y a pas lieu de nous y attarder bien longuement. Si des animaux adultes de même espèce et de même sexe ne sont pas identiques, c'est, d'une part, que leur constitution génétique présente des différences insoupçonnées et, d'autre part, qu'ils n'ont jamais vécu dans des conditions rigoureusement identiques. Il peut arriver que les différences de taille soient dues, pour la plus grande part, à un ou à un petit nombre de facteurs prédominants, certains gènes par exemple; mais on doit admettre qu'en général elles résultent de l'interférence d'innombrables actions, ayant chacune un effet très minime, qui se sont produites à un stade ou à un autre du développement, sous l'influence des conditions du moment. Si nous admettons que, pour la plupart au moins, ces causes de variation sont indépendantes et que leurs effets sont additifs, hypothèse que nous aurons à examiner de plus près dans un instant, il est possible, comme on sait, de démontrer que la distribution finale sera d'autant plus proche d'une distribution normale que le nombre des causes de variation sera plus grand et leur effet individuel plus petit. La démonstration s'étend au cas de distributions à deux variables, la corrélation s'expliquant, fait pour nous très important, par le fait qu'un grand nombre des causes qui font varier l'une des grandeurs agissent également sur l'autre et dans le même sens. Pour expliquer que certaines distributions s'éloignent nettement de la normale il suffit d'admettre que certains facteurs prépondérants du développement n'aient pas une distribution normale dans la population, que, par exemple, une partie des animaux ait été soumise, à un certain moment, à des conditions très différentes de celles où restaient placés les autres, ou encore qu'une partie des individus soit porteuse d'un gène modifiant les dimensions du corps tout entier ou celles de l'une de ses parties.<sup>2</sup>

Considérons alors deux races de taille moyenne inégale et supposons les si voisines l'une de l'autre dans leur forme et leur structure, que les seules différences existant apparemment entre elles soient d'ordre métrique; supposons les en outre également variables, au sens précis du terme, l'écart type étant proportionnel à la taille moyenne. D'après

---

<sup>2</sup> Dans ce cas on pourrait s'attendre à trouver des distributions de Pearson ou de Gram-Charlier.

ce qui précède, il faudra pour qu'il en soit ainsi que le nombre des causes perturbatrices, ou la grandeur de leur effet moyen, ou encore le nombre et la grandeur des perturbations élémentaires, augmentent avec la taille. Plus précisément, il faudra, dans les deux cas extrêmes, ou que, les perturbations conservant en moyenne la même importance, leur nombre soit sensiblement proportionnel à la racine carrée de la taille, ou que, ce nombre restant du même ordre, l'importance moyenne de chacune d'elles doit sensiblement proportionnelle à la taille. Il est clair que cette deuxième hypothèse est beaucoup plus raisonnable que la première et que l'on soit admettre que, lorsque la distribution est normale, les causes qui tendent à écarter un individu de la moyenne ont des effets sensiblement proportionnels à cette moyenne. Cette hypothèse dictée par des considérations théoriques a d'ailleurs un fondement biologique solide; elle signifie que les facteurs modificateurs de la croissance ont des effets multiplicatifs et non pas additifs, qu'ils agissent non pas en accroissant ou en diminuant la taille d'une certaine fraction de centimètre ou de millimètre, mais en l'accroissant ou en la diminuant d'une certaine fraction de cette taille. Ce fait paraît de plus en plus solidement établi pour les facteurs génétiques de la croissance et, en ce qui concerne les facteurs externes, toutes les interprétations physiologiques des lois de la croissance globale le postulent plus ou moins expressément.

Autant de raisons pour accepter cette hypothèse qu'il devient fort intéressant de pousser plus loin. Si l'effet d'un facteur apparaît comme multiplicatif lorsqu'on compare deux races de taille inégale d'une même espèce, on ne voit pas pourquoi il ne le serait pas aussi à l'intérieur d'une même race lorsque l'on compare les individus les plus grands et les plus petits, on ne comprend pas pourquoi l'effet d'un facteur sur un individu déterminé serait proportionnel à la taille moyenne de l'espèce et non pas à sa taille propre. On est ainsi conduit, par une généralisation toute naturelle, à interpréter la variabilité d'une population adulte par l'interférence d'innombrables actions élémentaires dont chacune augmente ou diminue d'un pourcentage très minime la taille de l'individu sur lequel elle s'exerce. Admettre qu'il en est ainsi revient à dire que la distribution des logarithmes des tailles doit être normale, ou que la distribution des tailles elles mêmes est du type de Galton-MacAlister. Les raisonnements précédents s'étendant immédiatement aux distributions à plusieurs variables, la régression logarithmique apparaît comme mieux fondée biologiquement que la régression linéaire.

Cet argument en faveur de la relation d'allométrie peut être présenté d'une autre façon qui a l'avantage de comporter moins d'hypothèses. On sait qu'une caractéristique des distributions normales à deux variables est d'être homoscédastiques, c'est-à-dire d'avoir, pour chaque variable, un écart type lié indépendant de la valeur particulière de l'autre. Autrement dit, si les individus sont classés selon les valeurs croissantes de  $x$ , l'écart type de  $y$  sera le même pour les individus les plus grands que pour les petits et, de même, si le classement est fait par rapport à  $y$ , l'écart type de  $x$  restera constant tout au long de la distribution. Si la distribution normale est celle des mesures directes, il s'ensuit que le coefficient de variabilité des grands individus sera inférieur à celui des petits. Si la distribution normale est celle des logarithmes des longueurs le coefficient de variabilité sera au contraire pratiquement indépendant de la taille. Cette remarque ne fournit pas pas moins un réel intérêt. On n'a jamais remarqué en effet que, dans chaque cas particulier, les individus de taille extrême étant généralement trop peu nombreux pour qu'un coefficient de variabilité puisse être déterminé sur eux avec quelque sécurité, mais elle n'en présente pas moins un réel intérêt. On n'a jamais remarqué en effet que, dans chaque espèce, les animaux les plus grands soient moins variables que les plus petits; tout porte à croire, au contraire, qu'il n'en est rien et cet argument en faveur de la relation d'allométrie mérite d'être retenu.<sup>3</sup>

Tenu compte de tout ce qui précède, nous opterons pour la relation d'allométrie contre la relation linéaire, en sachant bien cependant que ce choix raisonnable n'est pas obligatoire, puisqu'en matière de liaison stochastique, il reste toujours place pour un large arbitraire dans le choix des formes analytiques. Nous pourrions même dans certains cas user de cette latitude pour utiliser alternativement, selon le but poursuivi, l'une et l'autre des deux relations concurrentes.<sup>4</sup>

---

<sup>3</sup> On devrait constater par exemple que le coefficient de variabilité du propodite de la pince des *Maia squinado* mâles est de 9.8 pour les animaux mesurant 140 mm et de 4.1 pour ceux qui atteignent 220mm, avec une valeur voisine de 5.8 pour les animaux de taille moyenne 180mm, les changements de valeur de  $V$  dans un intervalle d'environ 2 écarts types de part et d'autre de la moyenne étant particulièrement importants dans ce cas ou la variabilité de l'organe considéré est très grande ( $V = 21$ ). Pour le méropodite du 4ème péréiopode dont la variabilité, encore, forte, est cependant plus normale ( $V = 12.7$ ) les coefficients de variabilité correspondant aux mêmes dimensions du céphalothorax devraient être respectivement 4.7; 2.8 et 3.5. Rien de tel n'a été observé.

<sup>4</sup> Voir en particulier l'usage que j'ai fait de cette possibilité dans une comparaison biométrique de deux espèces du genre *Maia*, *M. squinado* et *M. verrucosa* (*Comptes Rendus de l'Académie des Sciences*, T 204, p. 67, 1937).

Le seul argument qui pourrait incliner à préférer la traduction de la relation existant entre les deux variants par une formule linéaire plutôt que par une formule logarithmique est que, si deux parties d'un même organe comparées à une même grandeur de référence obéissent respectivement à des relations d'allométrie de constantes  $a_1$   $b_1$ ;  $a_2$   $b_2$ , il est impossible, si  $a_1$ , et  $a_2$  sont différents, que l'organe tout entier obéisse à une relation de la même forme. La difficulté ne se présente pas pour la relation linéaire où il suffit de poser :

$$\bar{y} = \bar{y}_1 + \bar{y}_2 \qquad a = a_1 + a_2$$

L'objection serait valable s'il s'agissait de relations fonctionnelles puisque dans ce cas, les seules relations répondant à la condition d'additivité doivent être de la forme  $y = a \phi(x) + b$ , condition que ne remplit pas la relation d'allométrie, ni aucune de celles que l'on pourrait songer à lui opposer, puisque la réciprocité entre  $y$  et  $x$  exigerait que l'on ait en même temps  $x = a' \phi(y) + b'$ , ce qui nous fait retomber sur la relation linéaire. J'ai montré, il y a plus de quinze ans, que cette objection est sans valeur dans le problème qui nous occupe, où la relation cherchée est stochastique. Sauf en des cas exceptionnels, on obtiendra une relation satisfaisante en prenant pour  $a$  la moyenne pondérée de  $a_1$  et  $a_2$  :

$$a = \frac{a_1 \bar{y}_1 + a_2 \bar{y}_2}{\bar{y}_1 + \bar{y}_2} = \frac{a_1 + a_2}{2} + \frac{(a_1 - a_2)(\bar{y}_1 - \bar{y}_2)}{\bar{y}_1 + \bar{y}_2}$$

La différence entre cette valeur moyenne de  $a$  et celle qui conviendrait pour l'extrémité de la distribution où l'écart est le plus grand est voisine de  $V_x (a_1 - a_2)^2$ . Elle est presque toujours dans la pratique inférieure à l'erreur que l'on peut commettre dans la détermination d'une constante d'allométrie, qui est, nous le verrons, de l'ordre de  $\alpha \sqrt{1 - r^2} / \sqrt{N}$ .

## II

Avant de poursuivre notre analyse de la notion d'allométrie, il nous faut traiter un problème fort important en lui-même et dont la solution est moins simple qu'on ne l'a cru. Une fois admis qu'une relation d'allométrie peut exister entre deux variables, il faut, pour vérifier cette hypothèse, disposer d'une méthode régulière qui permette de déterminer les constantes qui la définissent et la précision des chiffres

obtenus.<sup>5</sup> Une première question se pose immédiatement. L'ajustement doit-il être pratiqué sur l'expression  $y = bx^a$  ou sur l'expression équivalente  $\log y = a \log x + \log b$ . Les arguments donnés en faveur de l'ajustement de la première formule ne sont pas particulièrement convaincants. Les calculs sont d'une conduite plus difficile et donnent d'ordinaire, au surplus, des résultats fort peu différents de ceux que l'on obtient en utilisant la relation sous sa forme logarithmique. Comme par ailleurs nous avons conclu de l'étude précédente que la distribution de  $Y = \log y$  et  $X = \log x$  doit en principe être normale, il apparaît naturel de faire tous les calculs sur la deuxième expression que nous écrirons  $Y = aX + B$  et dans laquelle il s'agit de déterminer  $a$  et  $B$ .

La première solution qui se présente à l'esprit est celle qui consiste à identifier la droite cherchée avec la ligne de régression de  $Y$  en  $X$ . Dans le cas où nous nous sommes placés par hypothèse, celui de l'étude d'un échantillon représentatif de la population globale, cette solution ne peut pourtant pas être retenue. Elle suppose en effet qu'on est en droit de faire jouer un rôle différent aux deux variables, l'une  $X$  étant considérée comme indépendante et l'autre  $Y$  comme dépendante. Il peut évidemment arriver que cette hypothèse soit acceptable, si, par exemple  $Y$  est la mesure d'un organe de petite taille et  $X$  celle du corps tout entier, mais en général cependant,  $X$  et  $Y$  jouent des rôles symétriques. Si l'on doit comparer deux mesures linéaires du corps d'un animal, longueur et largeur du céphalothorax d'un crabe par exemple, ou de deux appendices homologues, on ne voit pas bien de quel droit on considérerait l'une plutôt que l'autre comme représentant la variable indépendante, ni par conséquent comment on pourrait choisir entre les deux lignes de régression. L'emploi de ces dernières ne se justifie que si, pour des raisons pratiques, on veut pouvoir calculer à partir de l'une des dimensions prise, par convention, comme référence, la valeur la plus probable de l'autre, mais dans ce cas on ne travaillera sans doute pas sur un échantillon représentatif de la population et l'on ne se trouvera donc pas dans le cas considéré actuellement. Cet emploi n'est généralement pas licite, les deux variables étant égales en dignité, à moins que l'on ne se résigne à donner au problème posé, non pas une,

<sup>5</sup> Le problème a été étudié en détail par Kavanagh et Richards (*Proceedings of the Rochester Academy of Science*, 150, 1942) qui ont montré que le traitement mathématique doit varier avec la nature des données de l'observation et le but poursuivi. Je renverrai à cet excellent travail pour tout ce qui touche à l'histoire de la question, à l'interprétation de  $b$  et à sa prétendue corrélation avec  $a$ . Ils ont justement insisté sur la nécessité qu'il y a dans certains cas, à faire jouer à  $y$  et  $x$  un rôle symétrique dans les calculs, mais sont restés dans le cadre des techniques traditionnelles.

mais deux solutions qu'il faudra toujours considérer simultanément. Si nous voulons arriver à une droite unique, la "meilleure" possible, nous devons renoncer au critère usuel de la qualité d'un ajustement et à l'emploi de la méthode des moindres carrés dans sa forme traditionnelle, rien n'autorisant à rendre minimum plutôt la somme des carrés des écarts des  $Y$  à  $X$  constant que celle des  $X$  à  $Y$  constant. Le problème est très voisin de celui qui se pose lorsqu'il s'agit d'ajuster une droite à un ensemble de couples de mesures où les deux variables sont sujettes à l'erreur, problème qui n'a été que peu travaillé jusqu'à une date récente et qui n'a pas encore reçu de solution vraiment satisfaisante. La meilleure paraît encore celle qui rend minimum la somme des carrés des distances des points à la droite; la droite obtenue se confond, on le sait, avec le grand axe des ellipses d'égale distribution. Du point de vue géométrique la solution est parfaitement raisonnable; du point de vue statistique elle l'est moins, car on ne voit pas très bien ce que peut représenter en terme d'erreur la distance d'un point figuratif à la droite.<sup>6</sup> Il m'a semblé qu'un autre principe de minimum pourrait être utilement envisagé et je propose de représenter la relation d'allométrie par la droite ( $D$ ) qui rend minimum la somme du produit des écarts de  $Y$  pour  $X$  constant et de  $X$  pour  $Y$  constant, ce qui revient à rendre minimum la somme des aires des triangles rectangles ayant pour hypoténuse commune la droite cherchée, deux côtés parallèles aux axes, et leurs sommets aux points figuratifs.

Un tel procédé n'est au fond que la transposition dans le domaine du calcul d'une technique d'interpolation graphique employée systématiquement par certains dessinateurs et que l'on utilise inconsciemment lorsque l'on essaie de placer au mieux une droite au travers d'un nuage de points approximativement alignés.

On voit sans peine que la droite cherchée ( $D$ ) passe par le centre de gravité  $\bar{X}$ ,  $\bar{Y}$ , de la distribution. Sa pente est donnée par la valeur de  $a$  qui rend minimum la somme des produits  $[(Y - \bar{Y}) - a(X - \bar{X})][(X - \bar{X}) - (Y - \bar{Y})/a]$ , c'est à dire par  $a^2 = \sigma_Y^2 / \sigma_X^2$ . La solution positive,  $a = \sigma_Y / \sigma_X$  donne la paramètre cherché, la solution négative,  $a' = -a$ ,

---

<sup>6</sup> Un autre inconvénient de la solution classique est que la droite obtenue dépend des unités adoptées pour la mesure des deux grandeurs comparées. Il suffit d'ailleurs pour y parer de normaliser les coordonnées, c'est à dire de prendre pour unités de mesure les écarts types des deux grandeurs. Mais si un tel artifice peut, à la rigueur, se défendre lorsque la comparaison porte sur les mesures mêmes, il ne saurait raisonnablement être utilisé lorsque les variables intervenant dans les calculs sont des logarithmes. Nous verrons cependant que la droite de moindre distance ainsi obtenue se confond avec celle à laquelle nous arrivons par des considérations très différentes et beaucoup plus acceptables.

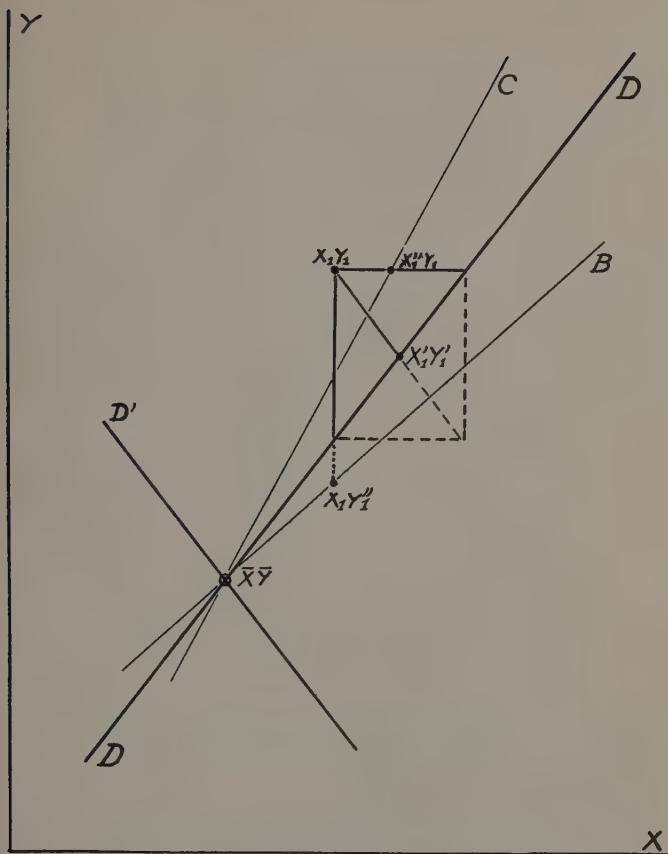


Fig. 3

Définitions comparées de la droite d'allométrie  $D$  et des deux droites de régression  $B$  et  $C$  et des points correspondant à  $X_1 Y_1$  sur  $D$  ( $X'_1 Y'_1$ ) et sur les droites de régression ( $X_1 Y''_1$  et  $X''_1 Y_1$ ).

est la pente de la droite ( $D'$ ) qui rend maximum cette somme de produits.

L'équation de la droite ( $D$ ) s'écrit :

$$Y - \bar{Y} = \frac{\sigma_Y}{\sigma_X} (X - \bar{X}) \quad \text{avec } a = \frac{\sigma_Y}{\sigma_X}$$

mais pourra aussi se mettre sous la forme symétrique :

$$\frac{Y - \bar{Y}}{\sigma_Y} = \frac{X - \bar{X}}{\sigma_X}$$

La variance de la constante d'équilibre  $a$  est  $\sigma a^2 = a^2(1 - r^2)/N$ ,  $r$  étant le coefficient de corrélation de  $X$  et de  $Y$ ; elle est égale à celle du coefficient de régression de  $Y$  en  $X$ . Les droites ( $D$ ) et ( $D'$ ) constituent un couple de directions conjuguées pour les ellipses d'égale distribution et le point  $X'_1, Y'_1$  de la droite ( $D$ ) correspondant aux mesures  $X_1, Y_1$ , relatives à un individu particulier est situé sur la parallèle à ( $D'$ ) menée par  $X_1, Y_1$ . Ses coordonnées sont :

$$X'_1 = X_1 + \frac{\sigma_X}{2} \left[ \frac{Y_1 - \bar{Y}}{\sigma_Y} - \frac{X_1 - \bar{X}}{\sigma_X} \right] \quad Y'_1 = Y_1 - \frac{\sigma_Y}{2} \left[ \frac{Y_1 - \bar{Y}}{\sigma_Y} - \frac{X_1 - \bar{X}}{\sigma_X} \right]$$

et l'on peut aussi écrire :

$$\frac{X'_1 - \bar{X}}{\sigma_X} = \frac{Y'_1 - \bar{Y}}{\sigma_Y} = \frac{1}{2} \left[ \frac{Y_1 - \bar{Y}}{\sigma_Y} + \frac{X_1 - \bar{X}}{\sigma_X} \right]$$

La variance des écarts entre les points figuratifs et les points de la droite ( $D$ ) qui leur correspondent est constante et égale à  $(\sigma_X^2 + \sigma_Y^2)(1 - r)/2$ . La mesure de la surface moyenne des triangles d'ajustement joue par ailleurs dans nos calculs le même rôle que la variance liée de la méthode classique de la régression et peut être qualifiée de covariance liée. Elle est proportionnelle à la grandeur précédente et constamment égale en tous points de la distribution à  $\sigma_X \sigma_Y (1 - r)$ . On voit que les formules qui précèdent permettent de faire correspondre sans ambiguïté à tout individu réel de la population l'individu normal fictif qui s'en rapproche le plus et de mesurer en écarts réduits leur degré de dissemblance qui peut être défini par l'expression :

$$\left[ \frac{Y_1 - \bar{Y}}{\sigma_Y} - \frac{X_1 - \bar{X}}{\sigma_X} \right] (1 - r)$$

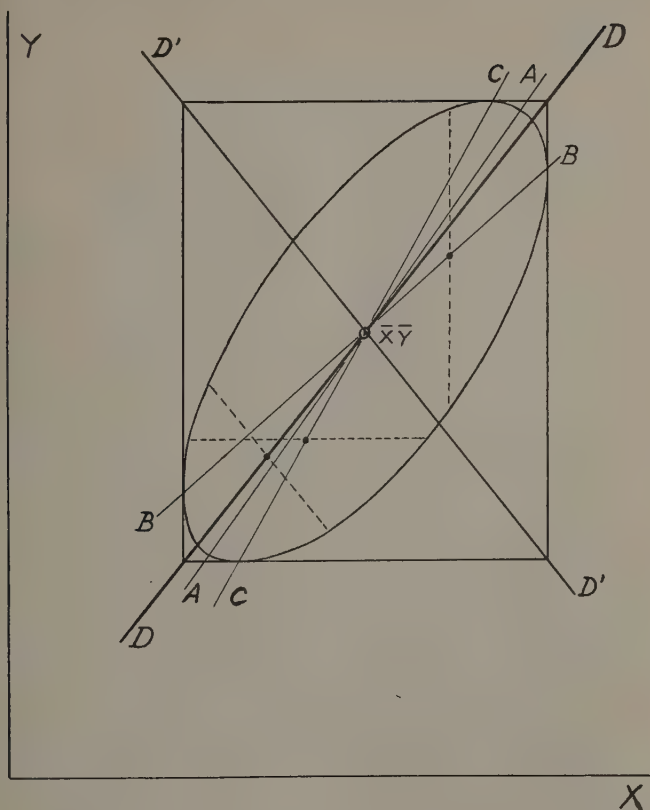


Fig. 4

Les droites  $D$  et  $D'$  mises en place dans une des ellipses d'égale distribution;  $B$  et  $C$ , les deux droites de régression et  $A$ , l'axe principal ( $\alpha = 1,285$ ,  $r = 0,71$ ).

L'usage qui vient d'être fait de la formule d'allométrie pour définir l'individu théorique correspondant à chacun des individus réels étudiés, est celui pour lequel elle a été établie et auquel elle est le plus exactement adaptée. Mais, dans la pratique, elle doit être utilisée beaucoup plus fréquemment à estimer la valeur de  $Y$  qui correspond normalement à une certaine valeur de  $X$  ou inversement à estimer la valeur de  $X$  correspondant normalement à une certaine valeur de  $Y$ .

Elle remplace ainsi, suivant le cas, l'une ou l'autre des deux lignes de régression de  $Y$  en  $X$  ou de  $X$  en  $Y$ .

L'estimation de  $Y$  qu'elle apporte diffère de celle que fournit la ligne de régression de  $Y$  en  $X$  de  $(1-r) \frac{\sigma_Y}{\sigma_X} \frac{X-\bar{X}}{\sigma_X}$ . Par ailleurs la variance de l'écart entre les  $Y$  observés et les  $Y$  calculés par cette formule pour une valeur donnée de  $X$  est :

$$\sigma_Y^2 (1-r^2) \left[ 1 + \frac{1-r}{1+r} \frac{(X-\bar{X})^2}{\sigma_X^2} \right]$$

que l'on peut aussi écrire :

$$2\sigma_Y^2 (1-r) + \sigma_Y^2 (1-r)^2 \left[ \frac{(X-\bar{X})^2}{\sigma_X^2} - 1 \right]$$

Elle dépend donc de  $(X-\bar{X})$ , sa valeur moyenne est  $2\sigma_Y^2 (1-r)$ . Elle n'est égale à la variance liée de  $Y$ , variance de l'écart entre les  $Y$  observés et les  $Y$  calculés par la ligne de régression, qui est égale, quel que soit  $X$ , à  $\sigma_Y^2 (1-r^2)$  qu'à sa valeur minimum atteinte pour  $X = \bar{X}$ . L'amplitude de ces variations est d'ailleurs faible. Il est exceptionnel que dans les échantillons de dimensions usuelles elle dépasse sensiblement, pour les individus les plus grands ou les plus petits, les erreurs commises normalement dans la détermination de la variance liée. Pour la même raison, et bien que les estimations fournies par la droite d'allogométrie présentent des écarts systématiques avec celles que permet la droite de régression, il est rare, lorsque la corrélation est forte, comme nous l'avons constamment supposé, que ces écarts soient notablement plus amples que ceux qui se rencontrent normalement entre points observés et points calculés par l'une ou l'autre droite. Il reste évidemment possible que, dans certains cas, et pour certains usages, il soit préférable d'utiliser la ligne de régression classique, mais cette réserve ne diminue en rien l'intérêt de la droite que nous avons définie.

Une des caractéristiques les plus remarquables de cette droite est que, parmi toutes celles qui pourraient être employées à décrire la liaison de  $X$  et de  $Y$ , elle est la seule dont la pente soit indépendante de la valeur du coefficient de corrélation de  $X$  et  $Y$ . Cette caractéristique, qui n'aurait peut-être pas grand intérêt dans l'étude d'une distribution quelconque, est fort avantageuse dans une étude de morphologie quantitative où, comme nous le verrons, la valeur numérique du coefficient de corrélation ne présente pas d'intérêt biologique. Il importe en

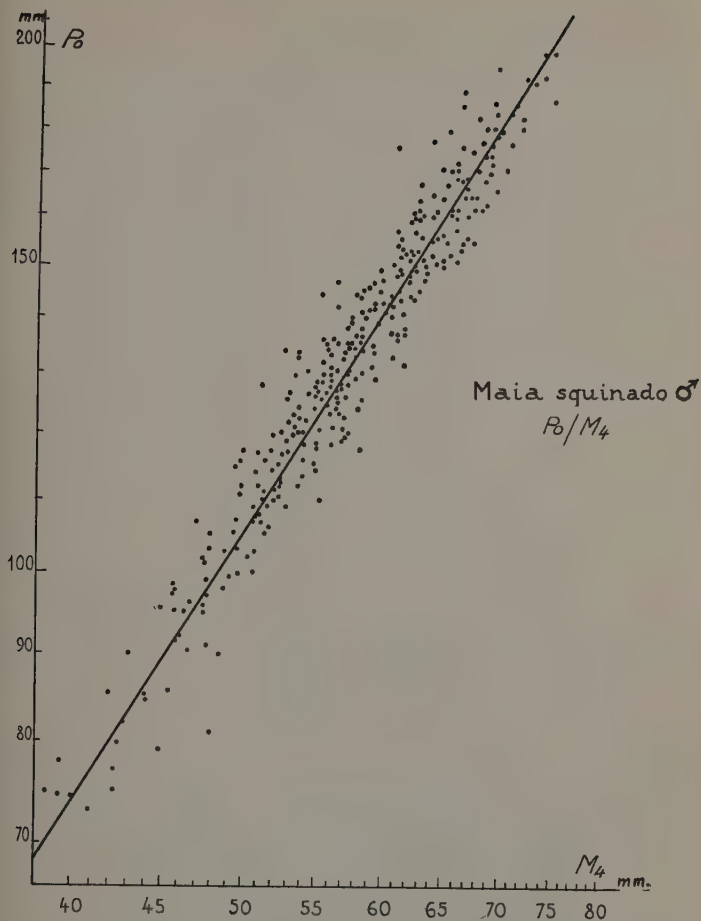


Fig. 5

Allométrie de taille chez les *Maia Squinados* mâles adultes. Longueur du propodite de la pince en fonction de la longueur du méropodite du dernier pereopode locomoteur chez 301 individus. Coordonnées logarithmiques ( $\alpha = 1,57$ ,  $r = 0,963$ ).

effet de bien comprendre que la loi que l'on essaie de dégager de l'ensemble des observations est, en définitive, celle que l'on pourrait raisonnablement espérer voir se manifester si la corrélation tendait vers la

perfection. Imaginons en effet une suite de distributions à deux variables à distributions marginales invariables et à corrélation croissante. On verra alors les lignes de régression se rapprocher l'une de l'autre et tendre, en même temps que l'axe principal vers une position limite qui est précisément celle de la droite (*D*) qui, elle, reste fixe.<sup>7</sup> C'est cette permanence qui, avec sa construction symétrique par rapport aux deux variables, lui donne pour nous un intérêt particulier, qu'accroît encore la facilité de son calcul.

Il importe cependant de préciser que ce calcul n'est valable qu'autant que le lot étudié constitue un échantillon représentatif de la population. Il cesserait de pouvoir être utilisé si les individus mesurés avaient été sélectionnés d'après les valeurs de l'une des variables. La seule méthode recommandable dans ce cas est la méthode des moindres carrés classique.

Un avantage important de la méthode régulière lorsqu'elle peut être employée et que l'on étudie simultanément plusieurs organes, est de permettre de donner en une fois toutes les relations d'allométrie qui peuvent exister entre toutes les variables prises deux à deux. Il suffit pour cela d'écrire les équations sous leur forme symétrique :

$$\frac{X - \bar{X}}{\sigma_X} = \frac{Y - \bar{Y}}{\sigma_Y} = \frac{Z - \bar{Z}}{\sigma_Z} = \dots$$

Toutes les relations cherchées se trouvent ainsi représentées, si l'on compare *p* organes, par une seule droite de l'espace à *p* dimensions, droite dont les diverses projections représentent les relations des organes pris deux à deux, trois à trois, etc. . . .

<sup>7</sup> La pente de la droite (*D*) est intermédiaire entre celle de l'axe principal et celle de la bissectrice des lignes de régression. Dès que la corrélation est forte ces trois droites sont très voisines. Avec  $r = 0.90$  et  $a = 0.900$  la pente de l'axe principal est 0.889 et celle de la bissectrice 0.9005 ; avec  $r = 0.95$  et  $a = 0.900$ , les pentes correspondantes sont 0.895 et 0.9001. La moyenne des pentes des deux droites de régression donne elle-même une très bonne approximation respectivement 0.905 et 0.901. Dans un cas comme celui des *Maia* ou toutes les corrélations dépassent 0.95, toutes ces droites se confondent pratiquement, les erreurs d'échantillonnage pouvant dépasser plusieurs millièmes. En intervertissant le rôle des variables dans l'exemple précédent, on verrait que les conclusions valables pour l'allométrie  $a = 0.900$  le sont également pour l'allométrie inverse  $a = 1.111$ .

Il est facile de voir que lorsqu'il y a isométrie ( $a = 1$ ), axe principal, bissectrices des lignes de régression et droite d'allométrie se confondent.

Nous obtiendrions la même superposition pour une valeur de *a* quelconque à la condition de faire usage de coordonnées normalisées. On peut donc dire, en un certain sens, que la solution proposée ici se confond avec celle qu'a préconisée Pearson, sans que la justification qu'il en avait donnée ait jamais paru pleinement satisfaisante. Il importe de noter à ce propos que le principe de minimum dont nous avons fait usage et qui nous paraît constituer la meilleure justification de la droite d'allométrie, ne semble avoir été invoqué, ni par Pearson, ni par aucun de ceux qui ont étudié plus récemment le même problème.

Sans vouloir développer ici cette extension des résultats précédemment obtenus, il est intéressant de noter que, si l'on prend en considération l'ensemble des  $p$  organes étudiés, on peut, comme nous l'avons fait pour deux organes seulement, faire correspondre à chaque individu réel,  $X, Y, Z, \dots$  un individu normal caractérisé par des mesures  $X'_1, Y'_1, Z'_1, \dots$  telles que :

$$\frac{X'_1 - \bar{X}}{\sigma_X} = \frac{Y'_1 - \bar{Y}}{\sigma_Y} = \frac{Z'_1 - \bar{Z}}{\sigma_Z} = \dots = \frac{1}{p} \left[ \frac{X_1 - \bar{X}}{\sigma_X} + \frac{Y_1 - \bar{Y}}{\sigma_Y} + \frac{Z_1 - \bar{Z}}{\sigma_Z} + \dots \right]$$

Tous les calculs précédents ont été faits sur les logarithmes et cette procédure est la seule qui soit recommandable. Des formules connues permettent d'en déduire les constantes qui caractérisent la distribution des mesures elles-mêmes. Dans la pratique les formules ci-dessous donnent une approximation toujours suffisante.

En appelant  $\bar{y}'$ , le nombre dont le logarithme est égal à la valeur armoyenne de  $\log y$ , on a approximativement :

$$\bar{y}' = e^{\bar{y}} \qquad \bar{y} = \bar{y}' (1 + \sigma_Y^2 / 2)$$

$$\sigma_y = \bar{y}' \sigma_Y (1 + 3\sigma_y^2 / 4) \qquad V_y = \sigma_Y (1 + \sigma_Y^2 / 4)$$

$\sigma_y$  ne peut guère dépasser 0.2, ce qui correspond déjà à un rapport de 1 à 4 entre les dimensions du variant pour les individus de taille extrême. On voit que l'on peut écrire à moins d'un centième près :

$$V_y = \sigma_y \qquad \sigma_Y = \bar{y}' \sigma_y$$

approximation toujours suffisante puisque l'erreur d'échantillonnage commise sur la détermination de l'un ou l'autre de ces paramètres peut dépasser largement le centième tant que l'effectif de l'échantillon est inférieur à quelques milliers. On aurait inversement, si les calculs avaient été faits sur les mesures mêmes :

$$\bar{Y} = \log \bar{y} - V_y^2 / 2 \qquad \sigma_Y = V_y (1 - V_y^2 / 2)$$

ou pratiquement :

$$\sigma_Y = V_y$$

Comme nous devions nous y attendre, la constante d'équilibre est le quotient des coefficients de variabilité des deux grandeurs comparées.

## III

Nous pouvons maintenant reprendre notre étude du problème biologique de l'allométrie.

Nous avons vu, dès le début de notre exposé, que l'allométrie de taille était très étroitement liée à la variabilité et nous venons de ramener à un seul ces deux problèmes. Définir la constante d'équilibre  $\alpha$  par le quotient des deux coefficients de variabilité revient à dire qu'il y a allométrie dès que les grandeurs que l'on compare sont inégalement variables et que l'isométrie ne peut exister que si les deux variabilités sont les mêmes. Nous avons interprété cette variabilité d'une population adulte par le jeu d'innombrables causes dont chacune a pour effet d'augmenter ou de diminuer d'un pourcentage très minime les dimensions d'un organe ou d'un autre. On sait, par ailleurs, et nous allons avoir l'occasion de le préciser, qu'il y aura corrélation entre deux organes si certains facteurs agissent simultanément sur l'un et sur l'autre, la corrélation étant d'autant plus complète que le nombre des facteurs est plus grand et leur action individuelle plus importante. Il est clair enfin que l'allométrie ne serait parfaite que si la corrélation l'était aussi, les variances liées ne devenant nulles qu'à cette occasion. Plus généralement, la relation d'allométrie sera d'autant plus exactement suivie que la corrélation entre les deux grandeurs comparées sera plus rigide et qu'il y aura moins de causes faisant varier un des organes indépendamment de l'autre.

Il serait par conséquent tout à fait déraisonnable de chercher une explication de l'allométrie de deux organes dans le jeu de certains facteurs qui agiraient sur l'un des organes et non pas sur l'autre, l'intervention de tels mécanismes ne pouvant, au contraire, que troubler la pureté du phénomène. S'il existe des facteurs de croissance agissant plus ou moins électivement sur telle ou telle partie de corps, leur action ne devra être invoquée que pour justifier l'existence d'écarts plus ou moins importants entre les prévisions permises par la relation d'allométrie et la réalité, l'allométrie même s'expliquant entièrement par l'inégale sensibilité des deux variants au même ensemble de facteurs. Si, par exemple, la longueur de propodite de la pince des *Maia* mâles comparée à la longueur du céphalothorax a une constante d'équilibre de 1.93, c'est qu'en moyenne le propodite de la pince des *Maia* est 1.93 fois plus sensible que le céphalothorax à toutes les actions qui peuvent s'exercer au cours du développement en le favorisant, ou le défavorisant.

Il est facile de préciser quantitativement ces indications en raison-

nant sur un modèle extrêmement simplifié, mais cependant fort instructif. Imaginons le cas idéalement simple où deux organes comparés réagiraient aux mêmes causes indépendantes de variation et à aucune autre, et où, pour chacune d'elles, l'effet serait proportionnel à l'intensité, le facteur de proportionnalité étant le même pour tous les facteurs et différent pour les deux organes. Adoptons pour chacun des facteurs un système de mesures tel que sa valeur moyenne soit nulle. Nous pourrions alors écrire,  $u_1, u_2, u_3 \dots$ , étant les facteurs de variation supposés très nombreux :

$$X = \overline{X} + \lambda u_1 + \lambda u_2 + \lambda u_3 + \dots + \lambda u_n$$

$$Y = \overline{Y} + \mu u_1 + \mu u_2 + \mu u_3 + \dots + \mu u_n$$

Nous savons que ces deux distributions sont normales ou approximativement normales et nous avons :

$$\sigma_x^2 = \lambda^2 (\sigma_{u_1}^2 + \sigma_{u_2}^2 + \sigma_{u_3}^2 + \dots + \sigma_{u_n}^2)$$

$$\sigma_y^2 = \mu^2 (\sigma_{u_1}^2 + \sigma_{u_2}^2 + \sigma_{u_3}^2 + \dots + \sigma_{u_n}^2)$$

La constante d'équilibre sera alors  $\alpha = \mu/\lambda$ , quotient des facteurs de proportionnalité. Le coefficient de corrélation sera égal à l'unité car :

$$r = \frac{\lambda \mu (\sigma_{u_1}^2 + \sigma_{u_2}^2 + \sigma_{u_3}^2 + \dots + \sigma_{u_n}^2)}{\sigma_x \sigma_y} = 1$$

Dans ce cas schématique, allométrie et corrélation seront simultanément parfaites, parce que toute la variabilité est conditionnée dans les deux organes par les mêmes causes.

Imaginons maintenant qu'on change dans les conditions d'existence de l'espèce, ou un changement génétique, fasse disparaître un groupe de causes de variabilité, mais leur en substitue deux autres, indépendants, qui exerceraient des actions de même nature et de même amplitude que les facteurs disparus, le premier agissant électivement sur l'un des organes, le second agissant électivement sur l'autre. Par hypothèse, les variances de  $X$  et  $Y$  restent les mêmes, mais le coefficient de corrélation diminuera, les nouveaux facteurs n'intervenant pas dans son numérateur, tandis que son dénominateur reste inchangé. Si  $v_1, v_2, v_3 \dots, w_1, w_2, w_3 \dots$  sont ces facteurs, qui exercent re-

spectivement, par hypothèse, des actions  $\lambda v_1, \lambda v_2, \lambda v_3 \dots$  sur  $X$  et  $\mu w_1, \mu w_2, \mu w_3 \dots$  sur  $Y$ , on trouve :

$$r = 1 - \frac{\lambda^2(\sigma_{v_1}^2 + \sigma_{v_2}^2 + \dots)}{\sigma_X^2} = 1 - \frac{\mu^2(\sigma_{w_1}^2 + \sigma_{w_2}^2 + \dots)}{\sigma_Y^2}$$

et :

$$2(1-r) = \frac{\lambda^2(\sigma_{v_1}^2 + \sigma_{v_2}^2 + \dots + \sigma_{w_1}^2 + \sigma_{w_2}^2 \dots)}{\sigma_X^2} = \frac{\mu^2(\sigma_{v_1}^2 + \sigma_{v_2}^2 + \dots + \sigma_{w_1}^2 + \sigma_{w_2}^2 \dots)}{\sigma_Y^2}$$

On vérifie aisément que les numérateurs de ces deux dernières expressions représentent les valeurs moyennes des variances liées de  $X$  et de  $Y$  ce qui nous redonne les expressions connues de ces grandeurs,  $2(1-r)\sigma_X^2$  et  $2(1-r)\sigma_Y^2$ .

On voit que, dans cet exemple, allométrie et corrélation sont indépendantes l'une de l'autre. L'allométrie reste constante parce que, par hypothèse, l'effet résultant de l'ensemble des facteurs reste le même, lorsque certains d'entre eux sont remplacés par d'autres; le coefficient de corrélation change avec le nombre des facteurs communs aux deux organes. Lorsque le nombre des facteurs communs diminue, la dispersion des points figuratifs autour de la droite d'allométrie augmente, ce que traduit l'accroissement des variances liées de  $X$  et de  $Y$ ; la détermination de  $a$  devient en même temps plus imprécise, comme le montre l'expression de sa variance,  $\alpha^2 (1-r^2)/N$ . La valeur de la constante d'équilibre  $a$  est ici égale à  $\mu/\lambda$ , c'est à dire au quotient des deux coefficients qui mesurent la sensibilité respective des deux organes aux actions perturbatrices. Nous voyons ainsi apparaître, avec une netteté toute particulière sur ce schéma simplifié à l'extrême, la caractéristique essentielle du phénomène d'allométrie que nous indiquions tout à l'heure à propos d'un cas concret : l'allométrie de taille exprime l'inégale variabilité de deux organes, témoignage elle-même de leur inégale sensibilité à l'ensemble des facteurs perturbateurs de la croissance.

Nous pourrions aisément compliquer le schéma précédent et nous rapprocher ainsi progressivement de la réalité. On conçoit sans qu'il soit besoin d'explication, qu'il peut arriver, qu'il doit même arriver, que les différents facteurs aient une importance inégale, que chacun d'eux ait ses coefficients d'action propres. Les résultats changeront naturellement, mais l'essentiel en subsistera. L'allométrie restera

imputable à l'ensemble des facteurs de la variabilité, la corrélation ne dépendra toujours que des seuls facteurs communs aux deux organes comparés; la variabilité résiduelle, que mesure la variance liée et qui est responsable des écarts des points observés et de la droite figurative, restera imputable aux facteurs qui agissent sur l'un ou sur l'autre des deux organes, mais non pas sur les deux.

Si notre étude se limite à la comparaison de deux organes, nous ne pouvons aller plus loin; mais si nous étudions plusieurs organes du même animal nous pouvons constater que chaque organe a sa variabilité propre et que, suivant les couples d'organes comparés, la corrélation est plus ou moins forte. On peut alors se demander si, en dehors des facteurs généraux qui agiraient sur toutes les parties de l'organisme et expliqueraient qu'entre deux organes quelconques existe toujours un certain degré de corrélation, n'existeraient pas d'autres facteurs agissant de façon plus ou moins élective sur la croissance de telle ou telle partie de l'animal. Ceux-ci pourraient d'ailleurs appartenir à plusieurs catégories, les uns étendant leur action à tout un système d'organes, les autres n'exerçant leur influence que sur une partie plus limitée du corps. Une telle étude, sans apporter aucun élément nouveau à la connaissance de l'allométrie même, permettrait de décélérer certaines influences qui l'empêchent de se manifester dans tous les cas avec la même rigueur. On peut espérer d'ailleurs qu'elle pourrait apporter quelque lumière sur des faits qui n'auraient même pas été soupçonnés sans elle.<sup>8</sup>

Posé dans ces termes, le problème est typiquement du ressort de l'analyse factorielle. Cette technique statistique, actuellement d'un emploi constant en psychologie, n'a encore été utilisée que très rarement en biométrie bien que, dès 1932, S. Wright en ait montré l'intérêt. Il faut espérer que les méthodes de calcul mises au point récemment par Delaporte en généraliseront l'emploi. J'en ai moi-même fait usage dans un travail consacré à l'étude des variants sexuels d'un Crustacé et j'indiquerai ci-après les principes qui m'ont guidé dans cette étude. Ils ne sont d'ailleurs que la généralisation de ceux que nous avons déjà mis en oeuvre.

---

<sup>8</sup> Je citerai à titre d'exemple, le problème de la taille cellulaire. Dans un travail consacré à l'étude des facteurs qui la conditionnent chez les Mammifères (*C. R. Soc. Biol.*, T. 135, 1941, p. 662, 750 et 1309), j'ai pu montrer en premier lieu que, d'une façon très générale, les dimensions du noyau sont unies à celles de la cellule par une relation de la forme  $N = bC^a$  ou  $b$  est une constante caractéristique du type cellulaire et ou  $a$  est voisin de 0.6 ou 0.7. J'ai établi ensuite que la taille moyenne d'une cellule dépend de deux facteurs, l'un général qui a des valeurs approximativement égales pour des animaux de même taille, quelle que soit leur espèce, l'autre spécial, qui ne dépend que de la nature de l'animal et non pas de sa grandeur.

Imaginons que nous ayons étudié un certain nombre d'organes et que les facteurs qui agissent sur eux se répartissent en un certain nombre de groupes indépendants entre eux  $u_1, u_2, \dots u'_1, u'_2, \dots u''_1, u''_2, \dots$ . Pour plus de commodité nous désignerons chaque groupe par une seule lettre, de sorte que nous pourrions écrire pour chaque organe :

$$Y_1 = \bar{Y}_1 + \lambda_1 U + \lambda'_1 U' + \lambda''_1 U'' + \dots$$

un ou plusieurs des coefficients  $\lambda_1, \lambda'_1, \lambda''_1, \dots$  pouvant être nuls. Le nombre des groupes de facteurs n'est pas défini a priori et peut varier avec le nombre d'organes étudiés, deux groupes pouvant agir simultanément sur plusieurs organes et séparément sur certains autres.

On aura alors :

$$\sigma_{Y_1}^2 = \lambda_1^2 \sigma_U^2 + \lambda'^2_1 \sigma_{U'}^2 + \lambda''^2_1 \sigma_{U''}^2 + \dots$$

la fraction de la variance totale attribuable à  $U$  est  $\tau_1^2 = \lambda_1^2 \sigma_U^2 / \sigma_{Y_1}^2$  et  $\tau_1$  représente le coefficient de corrélation  $r_{1U}$  de  $Y_1$  et de  $U$ . On montre sans difficulté que le coefficient de corrélation entre deux organes  $Y_p$  et  $Y_g$  est :

$$r_{pg} = r_{pU} r_{gU} + r_{pU'} r_{gU'} + r_{pU''} r_{gU''} + \dots$$

Les coefficients de corrélation  $r_{rg}$  pouvant être mesurés directement il s'agit de calculer à partir d'eux les coefficients  $r_{pU}, r_{pU'}, \dots r_{gU}, r_{gU'}$ . Il n'existe pas de méthode générale de résolution d'un tel système d'équations et l'on doit procéder à une série de tâtonnements méthodiques, en essayant successivement des schémas à 1, 2, 3 . . facteurs, jusqu'à ce que l'on arrive à retrouver les  $r_{pU}$  avec les  $r_{pg}$ .

L'application de cette méthode<sup>9</sup> m'a permis de reconnaître chez les *Maia* un *facteur général* qui conditionne la taille globale, évaluée par les dimensions du céphalothorax, dont la mesure  $L$  est la meilleure estimation et qui contrôle 92 à 94% de la variance totale des appendices; un *facteur de groupe*  $G$  qui agit sur l'ensemble des appendices et contrôle 3 à 6% de leur variance; un *facteur régional*  $R$  qui agit électivement sur les appendices antérieurs dont il contrôle 3% environ de la variance. La variance résiduelle, qui est de l'ordre de 1% de la variance totale, est imputable à des facteurs locaux et ne peut être évaluée que par différence.

Il est possible à partir de l'ensemble des mesures des appendices de calculer les valeurs numériques de  $G$  et de  $R$  relatives à chaque individu,

<sup>9</sup> Voir *Biotypologie*, T. 6, pp. 73-98, 1938.

et d'établir, d'autre part, des formules de prévision plus complètes que la relation d'allométrie, en ce qu'elles font intervenir simultanément  $L$ ,  $G$ ,  $R$ . On constate alors qu'il est possible de recalculer, à partir de cet ensemble de trois nombres, les dimensions de chaque individu avec une précision supérieure au centième, les coefficients de variabilité résiduelle étant tous inférieurs au dixième des coefficients de variabilité initiaux.

On ne peut pas prévoir, tant que d'autres recherches sur ce problème n'auront pas été faites, quel peut-être l'avenir de ce genre d'étude. Il n'est pas interdit d'espérer qu'elles pourront apporter quelques lumières sur les où la relation d'allométrie se vérifie imparfaitement, soit que le nuage de points figuratifs soit trop étalé, soit qu'il s'incurve à l'une de ses extrémités.

#### IV

Il s'agit de montrer maintenant que le problème de l'allométrie de croissance n'est pas différent au fond de celui de l'allométrie de taille.

Nous remarquerons en premier lieu que rien de ce qui a été dit plus haut de l'allométrie de taille n'implique que les animaux comparés aient définitivement cessé de croître, mais que nous avons seulement supposé que la population étudiée ne comportait que des individus ayant atteint le même stade de développement. Nos conclusions vaudraient, par exemple, tout autant pour chacun des stades larvaires d'un Insecte que pour son stade imaginal. C'est à partir de cette très simple remarque que nous allons pouvoir construire la démonstration cherchée que nous établirons tout d'abord sur le développement d'un Arthropode, la discontinuité de la croissance dans cet embranchement rendant les faits particulièrement clairs.

Considérons les stades successifs du développement d'un de ces animaux, un Crustacé par exemple, en les choisissant dans une même phase de la croissance, c'est à dire dans une période de l'existence où ne se produit aucun remaniement organique, ni aucun changement notable dans le fonctionnement physiologique des différents appareils. Supposons, plus précisément, que nous sommes dans les conditions où une étude de la croissance relative a un sens, c'est à dire que la forme d'un individu ne dépend que de sa taille. Il peut arriver que la population de chacun de ces stades soit assez peu variable et que le taux d'accroissement lors de la mue soit assez grand pour que les représentants de deux stades consécutifs aient des dimensions qui ne se chevauchent pas. En général cependant les individus les plus grands du premier stade ont des dimensions supérieures à celles des individus les plus petits du second stade. Autrement dit, les nuages figuratifs

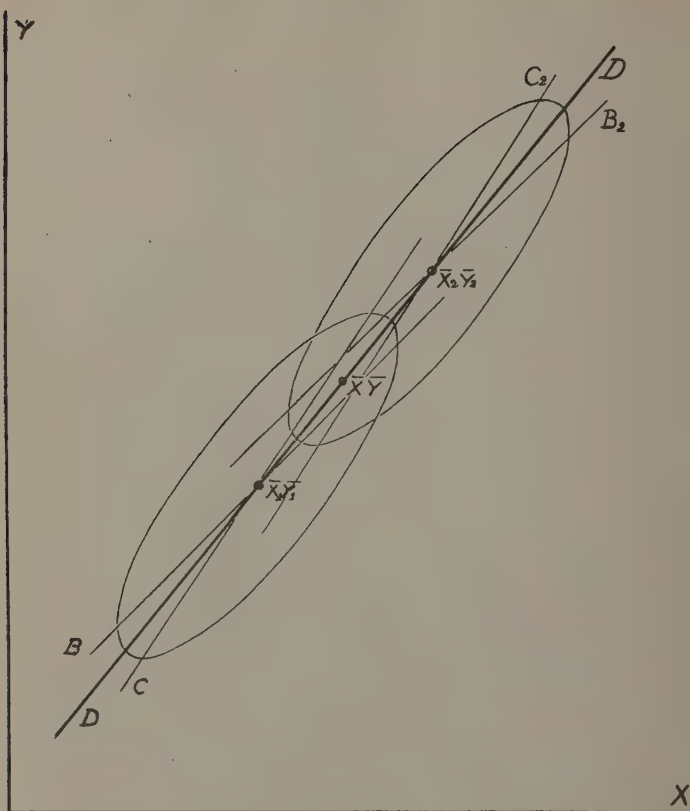


Fig. 6

Schéma correspondant à celui de la fig. 4, dans l'hypothèse où la population comporte en nombre égal les représentants de deux stades consécutifs du développement. Voir le texte.

des deux populations de mesures de  $x$  et de  $y$  se recouvrent partiellement. Comme nous avons supposé expressément que dans l'espèce étudiée la forme ne dépend pas directement de l'âge, les deux séries de mesures de la zone de recouvrement doivent nécessairement correspondre à la même loi de croissance et donner des résultats aussi concordants que si elles avaient été pratiquées dans la même zone sur deux

populations appartenant à un même stade. Il résulte de là que, si une courbe puissance traduit dans cette zone pour la première population la relation existant entre les deux variants, elle doit la traduire tout aussi bien pour la deuxième. Les relations d'allométrie  $f$  correspondant à la première population et  $f_2$  correspondant à la deuxième, qui coïncident dans la zone de chevauchement, coïncideront alors dans toute leur étendue. Il en serait de même pour tous les stades qui précèdent ou qui suivent les deux stades considérés jusqu'à présent, à la seule condition que dans toute cette étape du développement, on soit en droit de considérer la forme comme fonction de la taille. La loi de croissance  $f$  sera en ravanche remplacée par une loi différente  $\phi$  lorsque le développement sera arrivé à un stade tel que, dans la zone de recouvrement, les deux populations doivent être considérées comme différentes.

Le problème de la recherche d'une loi de croissance relative valable pour une des étapes du développement d'un Arthropode se ramène ainsi, dans les hypothèses où nous sommes placés, à celui de la recherche de la relation existant entre deux organes pour les animaux d'un même stade choisi arbitrairement dans cette étape. Ainsi s'explique, de la façon la plus naturelle, l'identité formelle des relations qui permettent de comparer les formes successives réalisées au cours de développement, ou les formes diverses que peuvent présenter suivant leur taille les animaux d'un même stade. Ainsi s'explique aussi très simplement, le fait que le passage d'une étape à la suivante se marque sur la courbe de croissance relative par un point anguleux ou une discontinuité, puisqu'il n'y a aucune raison pour que les courbes  $f$  et  $\phi$  se raccordent tangentiellement l'une à l'autre.

L'élément nouveau que nous venons d'introduire dans la discussion va nous permettre de donner au problème de la détermination des constantes de la courbe une autre présentation que celle que nous connaissons déjà. Il faut que le procédé de calcul de ces constantes soit tel qu'il donne le même résultat, lorsqu'il porte sur les mesures correspondant à un même stade, et lorsqu'il est appliqué à deux ou à plusieurs stades successifs. Pour nous placer dans un cadre plus concret, nous adopterons un mode de transcription graphique des résultats qui rende sensiblement normal le nuage des points figuratifs de la population des mesures relatives à un stade donné. Nous savons que la résultat sera atteint fréquemment en utilisant les coordonnées logarithmiques; parfois pourront être employées tout aussi bien des coordonnées arithmétiques. La corrélation entre les deux variants étant forte par hypothèse, le nuage sera très allongé suivant une droite

sur laquelle s'échelonneront aussi, d'après ce qui vient d'être dit, les nuages figuratifs des autres stades. La définition des paramètres de cette droite doit être telle que le résultat obtenu soit le même, que l'on ait utilisé un seul des nuages de points, plusieurs d'entre eux, ou leur ensemble. Un calcul très simple montre que cette condition suffit à exclure les droites de régression, que nous savions déjà ne répondre que très imparfaitement au problème posé, et, au moins lorsque les coordonnées sont logarithmiques, l'axe principal qui, fournissant une solution symétrique en  $X$  et  $Y$ , aurait pu à la rigueur être retenu. Il est clair, en revanche, que si le rapport des variances de  $Y$  et  $X$  a même valeur pour toutes les populations, ce que nous avons supposé, en admettant que l'allométrie est la même pour deux stades consécutifs, le résultat cherché sera atteint en prenant pour pente de la droite le quotient des écarts types des variables comparées qui, si les nuages de points sont réellement alignés, est le même pour chacun d'eux et pour le nuage global. Nous retrouverons ainsi, par une toute autre voie, le résultat auquel nous étions arrivé en étudiant l'allométrie de taille.

Ainsi se trouve établie, pour un Arthropode, la proposition annoncée, mais avant de poursuivre, il importe de préciser que les conclusions auxquelles nous sommes arrivés offrent un degré assez inégal de généralité.

Nous avons démontré qu'à l'intérieur d'une étape de la croissance une même loi  $f$  convient à la fois à l'allométrie de taille pour chacune des stades et à l'allométrie de croissance pour l'ensemble de ces derniers. Nous avons démontré accessoirement que deux étapes successives sont normalement séparées sur la courbe figurative du phénomène par un point anguleux ou une discontinuité, marques du passage d'une loi de croissance à une autre. Mais nous n'avons pas démontré qu'il existait de telles étapes et ne pouvions le démontrer puisqu'il s'agit là d'une question de fait. L'observation montre d'ailleurs que s'il existe de nombreux cas où une étape comporte cinq ou six stades ou même davantage, il en est d'autres, tels que les développements postpubéraux de certains Crabs, où chaque stade doit être considéré comme constituant à lui seul une étape distincte. La courbe totale peut ainsi comporter à son extrémité un ou plusieurs segments ne représentant qu'une allométrie de taille, alors que les précédents décrivent à la fois une allométrie de croissance et une allométrie de taille. C'est le cas pour *Maia squinado*, où les courbes traduisant le développement des appendices dans les deux sexes, ou celui de l'abdomen des femelles, com-

portent dans la partie actuellement connue, trois segments, dont le dernier, séparé du précédent par une discontinuité, se rapporte exclusivement à l'allométrie de taille de l'adulte.

Plusieurs hypothèses supplémentaires ont été nécessaires pour que puisse être précisée la forme de la relation  $f(x, y) = 0$ . Nous avons supposé que nous disposions d'un mode de transcription de nos résultats qui rendait sensiblement normal, à chaque stade, le nuage de nos points figuratifs. La démonstration précédente implique bien que, si ce résultat peut être atteint pour l'un des stades, il l'est nécessairement

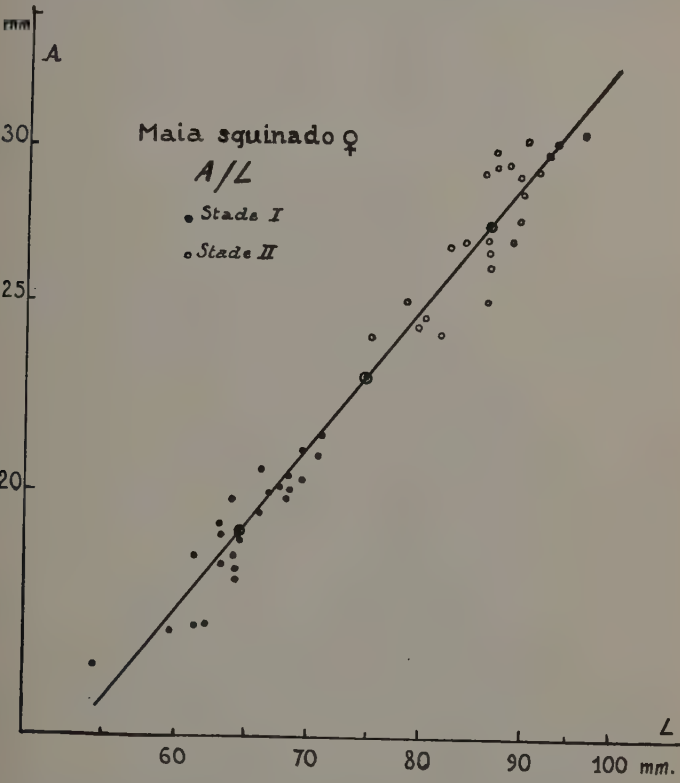


Fig. 7

Allométrie de l'abdomen chez les *Maia squinados* femelle, à deux stades consécutifs du développement ( $\alpha = 1,28$ ).

pour tous ceux qui appartiennent à la même étape, puisque la relation qui unit les deux variables est unique et que devenue linéaire dans une de ses parties, elle le devient nécessairement dans les autres. Mais elle ne peut pas nous dire quel est le type d'anamorphose qui convient à chaque cas. S'il est vrai que le résultat peut être souvent atteint en utilisant des coordonnées logarithmiques, et si nous avons de bonnes raisons de croire qu'il en est très généralement ainsi, nous connaissons quelques cas où il en va autrement. Ce sont ceux qui sont connus sous le nom d'allométrie variable et où la courbe logarithmique présente une convavité plus ou moins marquée ou même un point d'inflexion. Ils s'observent surtout, semble-t-il jusqu'à présent, sur les tronçons de la courbe totale qui se rapportent aux adultes et représentent essentiellement des allométries de taille. Mais il n'est pas exclu que l'on puisse trouver quelque jour des allométries de croissance du même type.

Enfin nous rappellerons que le choix de la "meilleure courbe" représentative d'un ensemble de mesures comporte toujours une large part d'arbitraire et que la solution à laquelle nous sommes arrivés, pour satisfaisante qu'elle puisse être, n'est pas la seule que l'on soit en droit de retenir. Nous préciserons également que cette solution est essentiellement théorique, qu'elle n'est valable que lorsque sont remplies au préalable toute une série de conditions et qu'au surplus elle ne prétend, en aucune manière, à être une méthode régulière de calcul des constantes d'une courbe d'interpolation.

Les remarques et les réserves qui viennent d'être faites, pour utiles qu'elles puissent être, ne doivent cependant pas être considérées comme diminuant la portée de notre démonstration. Elles en précisent seulement la signification et rappellent, s'il en était besoin, qu'un schéma théorique ne doit pas être appliqué à un cas concret sans d'indispensables précautions. Il nous reste à montrer rapidement que les résultats précédents s'étendent aux organismes dont la croissance ne se fait pas par mues. Si nous avons choisi, pour notre démonstration, le cas des Arthropodes, c'est parce que, dans cet embranchement, les divers stades définissent clairement une succession de populations sur chacune desquelles il est possible d'étudier les phénomènes d'allométrie de taille. Mais il est clair qu'une telle étude peut également être faite sur des animaux à croissance continue, si l'on possède un critère permettant de définir sans ambiguïté les individus comparables, critère qui évidemment pourra la cas échéant être l'âge. Les raisonnements

précédents restent valables, sous la réserve habituelle que, dans la période étudiée n'intervienne pas de phénomène physiologique capable de donner à deux animaux de même taille des structures dissemblables. Pour le reste, la différence intéressante entre la cas général et celui des Arthropodes porte sur la définition des stades successifs qui, au lieu d'être imposée par la nature même du développement, peut devenir assez largement arbitraire. On peut, en principe, multiplier ces stades autant qu'on le désire, sans que rien soit changé à nos raisonnements, ce qui permet d'étendre la démonstration du cas des animaux à développement discontinu à celui des animaux à croissance continue.

Quelques lignes nous suffiront pour étendre à la comparaison biochimique des organismes les procédés employés dans la comparaison morphologique. Rien, dans les formules que nous avons utilisées, n'implique que  $x$  ou  $y$  doivent mesurer plutôt la longueur d'un appendice que le poids du cerveau ou la quantité de calcium renfermée dans un organisme. Toutes les interprétations valables dans un cas le sont aussi dans l'autre et la notion d'allométrie biochimique, qui a été vivement critiquée, mais qui a aussi été employée avec un grand succès, sans jamais pouvoir être expliquée de façon satisfaisante, apparaît maintenant comme pleinement justifiée.

\* \* \*

Les nombreuses recherches de morphologie quantitative de ces vingt dernières années ont permis de découvrir un grand nombre de faits nouveaux et de donner une forme précise à des notions dont les travaux antérieurs avaient pu faire pressentir l'importance. Ces recherches n'ont mis le plus souvent en oeuvre que des méthodes très élémentaires et un appareil mathématique très réduit.

La simplicité de ce mode d'étude a beaucoup facilité sa diffusion ; il a permis d'acquérir en assez peu de temps des renseignements précieux sur la croissance d'animaux très divers et d'établir des comparaisons précises entre adultes de même espèce ou d'espèces voisines. Il est certain que des recherches de ce genre conservent à l'heure actuelle une bonne part de leur utilité et que, longtemps encore, elles apporteront, dans certains domaines, des contributions très appréciables à nos connaissances.

Mais, dans un nombre croissant de circonstances, il apparaît de plus en plus clairement que des conclusions légitimes ne pourraient être tirées de telles études qu'à la condition de disposer de critères plus sûrs que ceux que permet le simple ajustement graphique. Aussi

a-t-on été conduit à joindre à la technique rapide et facile, mais seulement approximative, de l'allométrie, les techniques plus laborieuses et plus savantes, mais aussi plus rigoureuses, de la biométrie classique. De plus en plus fréquemment est apparue, au côté des constantes de la formule d'allométrie, l'indication de leurs écarts type. Ce faisant, on se contente d'ailleurs de juxtaposer l'emploi de deux procédés, l'un d'origine biologique, l'autre d'origine mathématique, sans les fondre en une technique cohérente.

La notion même d'allométrie, malgré toutes les tentatives qui ont été faites pour en pénétrer la signification biologique, n'est encore aujourd'hui guère plus qu'une simple donnée de l'expérience, dont le fondement rationnel reste mal assuré, et que beaucoup de ses partisans considèrent toujours comme essentiellement empirique. La diversité même des circonstances dans lesquelles il paraît permis d'en faire usage en rend la compréhension plus obscure et la contradiction s'aggrave entre l'importance grandissante des résultats obtenus et la précarité de plus en plus apparente de leur fondement. Pour surmonter cette contradiction, il est devenu nécessaire de repenser la notion même d'allométrie.

C'est ce que nous avons essayé de faire dans les pages qui précèdent. Éliminant les difficultés accessoires, négligeant provisoirement tous les phénomènes qui, plus ou moins légitimement, apparaissent en désaccord avec la simplicité de la relation fondamentale, nous avons essayé de comprendre la signification véritable de la relation d'allométrie. Il m'apparaît que nous y avons à peu près réussi et qu'elle n'est rien d'autre qu'un aspect de la variabilité essentielle de tous les êtres vivants.

## DISCUSSION

**Lois M. Zucker.** I should like to discuss some data we have which fit in rather well with a question Dr. Teissier has raised. We have data on the relations between the femur ash and body weight in growing rats, selected in five age groups. Each age group provides a scatter diagram relating log-ash to log-weight and these scatter diagrams overlap as in Dr. Teissier's example. Now, does the same line or the same slope apply both within ages and between ages? In other words is the relation independent of age? The choice of line to use in fitting the data

becomes rather critical. Neither variable is independent and there is error in both variables. Therefore the preferred line may fall between the two regression lines. If the deviations parallel to log-ash are minimized the five slopes do not differ significantly from a common trend which, in turn, is very much flatter than the trend between ages; if the deviations parallel to log-weight are minimized, the five slopes have a common trend which does not differ significantly from the trend between ages.

The line suggested by Dr. Teissier lies somewhere between. But is it the best line? It seems to us that it has a flaw. Consider the ideal line expressing the relationship. Suppose there are only, say, ten possible type positions along this line instead of a continuous set. Neither variable is independent. Actual points will range themselves around their type position in elliptical distributions which, accordingly to the error distribution, may be nearly circular (each variable with equal error) or tall and narrow (most of the error with the  $Y$  coordinate) or low and wide (most of the error with  $X$ ). In the last case the regression of  $X$  on  $Y$  is a good estimate of the true line while the regression of  $Y$  on  $X$ , and to a lesser extent lines like the one suggested by Dr. Teissier which he approximated midway between the two regression lines, are too flat in slope. In the case with most of the error with the  $Y$  coordinate, the regression of  $Y$  on  $X$  is a good estimate, where the regression of  $X$  on  $Y$ , and to a lesser extent Dr. Teissier's line, are too steep. It is only if the error is approximately the same for both variables that a line between the two regression lines, such as that suggested by Dr. Teissier is a good estimate. It does not seem possible to accept a single solution independent of the error distributions for all these situations, like Dr. Teissier's line or the one suggested by Dr. Wald in a recent publication.

In response to Dr. Teissier's suggestion that his line is recommended only when neither variable is independent, I should like to state that a scatter diagram like that obtained with one variable independent can also arise for neither variable independent, if the error is mostly associated with one variable. It is usually impossible to estimate objectively the relative error associated with the two variables in cases like this, because the error is largely biological and not error of measurement.

We believe, for instance, that in the relation between femur ash and body weight much more error is associated with body weight than with ash. It is very difficult to affect the ash by experimental means—water deprivation, starvation, disease, or nutritional deficiencies other than of calcium, phosphorus or vitamin D, while the body weight, largely

affected by the weight of water and soft tissues, is very sensitive to experimental procedure. Should it not, therefore, be much more variable? For this situation we should prefer a line very close to the regression of log-body weight on log-femur ash, not a line halfway towards the other regression line. However, our belief as to the error distribution cannot be directly verified objectively.

(For data see Zucker and Zucker, *American Journal of Physiology*, 146, 585, 593, 1946; for further discussion see Zucker, *Human Biology*, 19: 232, 1947.)

**Harold Hotelling.** From a purely descriptive standpoint, data on a pair of variates such as measures of a bone cannot possibly be summarized adequately by any one regression line. Even if we make the simplifying assumption of a bivariate normal distribution, there are five parameters, and these cannot be summarized by the two coefficients in the equation of a straight line.

When we pass over to the relation of bone dimensions or chemical composition to vitamin D in the diet we are on different ground. If the object is simply to predict bone ash percentage as a function of vitamin D, then a simple regression equation for bone ash percentage on vitamin D quantity is in order, and this equation will disregard all measures of the bone excepting ash content. If, however, there is a question whether vitamin D has some effect on the bone, without advance knowledge of the particular kind of effect, and if say eight measures on each bone are taken to represent it fully, then a more general method should be used. This is arithmetically identical with the technique of fitting a multiple regression equation and testing the significance of  $R$ .

The question at issue may be the still broader one of finding some combination of vitamins that will affect in some way, not initially known, the shape or composition of a bone. Then the appropriate method is to determine together the most predictable function of the bone dimensions and contents, and the best function of the various vitamin doses for predicting this. Techniques for doing this, and for testing the significance of the results, have been developed (*Biometrika*, 28: 321-377, 1936). This kind of approach will, I think, supersede, for this kind of purpose, the summarization of the relations between measures on a bone by any regression line or its  $n$ -dimensional generalizations.

**N. Rashevsky.** The statistical relations discovered by Professor Teissier constitute important new knowledge. This kind of knowledge concerns how phenomena happen. If we wish the answer to the question of why they happen, we must extend our approach by introducing hypotheses and postulates. Professor Teissier is well aware of this himself, since he mentioned that he had a physiological explanation of some of these phenomena.

To illustrate what the theoretical method can do with the problem of growth, I wish to mention some studies in the Section of Mathematical Biophysics at the University of Chicago, relating to the problem of cancer.

Multicellular organisms grow up to a certain limit, whereas in tissue cultures cellular multiplication may go on practically indefinitely. This suggests that some inhibitory factor is produced by each cell, so that as the total number of cells of the organism increases and each cell becomes inhibited by a greater and greater number of cells, growth eventually stops. The mathematical formulation of such an assumption leads to growth curves reminding us of those described by Professor Teissier. As has recently been shown by Kesselman, some aspects of allometric growth can be described similarly. The theory implies that after the adult stage is reached, the total amount of the inhibitory factor decreases due to the repair of the natural wear of the organism and of accidental wounds. The theory leads to an expression for the decrease of velocity of wound-healing with age. Although the data on this question are very meager, whatever is available is in agreement with the theory. As the organism grows older and the total inhibitory effect decreases, small accidental fluctuations will be sufficient to produce accidental growth. Hence such accidental growth would be expected to be more frequent in older age. Actually, it is known that the incidence of cancer does increase with age.

The theory also enables us to derive the incidence curve as a function of age and the theoretical curve is found in good agreement with the observed one. From one of the parameters of the incidence curve we can compute the natural human longevity, longevity as it would be if unaffected by illness or accident. The theoretically computed value turns out to be of the order of 80 to 100 years. It is certainly remarkable that from a single theory we may compute such different things as decay in wound-healing ability, incidence of cancer and human longevity. But the most important progress in science consists principally in finding relations between phenomena which at first sight appear to

be unrelated. In cancer research the purely statistical method is of extreme importance to give us the data. But for a complete success it must be supplemented by a broader theoretical approach.

**Georges Teissier:** J'ai suivi avec intérêt les remarques présentées à la suite de mon exposé, mais si je crois avoir compris le sens général de ces interventions, je ne suis pas assez assuré d'en avoir suffisamment saisi les nuances pour pouvoir les discuter en détail.

A. Mme. Lois ZUCKER je répondrai seulement qu'il ne me semble pas que nous parlions exactement des mêmes choses. Les hypothèses dans lesquelles elle se place, pour les deux courbes qu'elle étudie, diffèrent trop des miennes pour que des conclusions valables dans le cas dont je me suis occupé puissent s'appliquer sans modification aux schémas qu'elle a envisagés.

Au Dr. HOTELLING, je répondrai que je sais très bien que la ligne d'allométrie, non plus qu'aucune ligne de régression d'aucune espèce ne peut fournir à elle seule la description complète d'un ensemble de mesures. Je n'entends pas substituer à la description statistique classique des faits un procédé plus simple et aussi efficace, mais bien fournir de l'essentiel de ces faits, un schéma clair et interprétable en termes biologiques. L'immense littérature consacrée depuis plus de vingt ans à la croissance relative établit sans conteste que la relation d'allométrie remplit parfaitement ce but. Je me proposais au jourd'hui, non pas d'établir sa validité pratique, qui est certaine, mais seulement de dégager sa double signification, biologique et statistique, résultat auquel je crois bien être parvenu.

À M. RASHEVSKY je dirai que je suis moi-même responsable d'un schéma physiologique qui permet, à ce que je crois, de justifier de façon très suffisante l'application de la relation d'allométrie à un organisme en croissance, et qui explique en même temps l'existence des stades critiques et des étapes du développement. Mais ce schéma ne vaut ni pour l'allométrie de taille, ni pour l'allométrie biochimique.

Comme il serait déraisonnable de supposer que l'identité des relations qui interviennent dans les trois cas puisse résulter d'une simple coïncidence et que la justification d'une relation applicable en des circonstances aussi dissemblables puisse reposer sur des considérations mécanistes, il faut chercher dans une autre voie. Je crois avoir réussi à montrer qu'une interprétation statistique répond entièrement à la question posée. Cela ne signifie naturellement pas qu'il faille renoncer à connaître les mécanismes physiologiques et physicochimiques qui interviennent dans chaque cas. Tout en étant décidé, pour ma part, à continuer l'étude que j'ai entreprise de cet aspect du problème, je tiens à déclarer qu'à mon avis, et tenu compte d'innombrables tentatives dont celle de Robertson reste le prototype, il est dangereux de vouloir fonder une théorie de la croissance sur un schéma physico-chimique nécessairement trop simple.

# THE GENERAL THEORY OF PRIME-POWER LATTICE DESIGNS\*

## I. INTRODUCTION AND DESIGNS FOR $p^n$ VARIETIES IN BLOCKS OF $p$ PLOTS

OSCAR KEMPTHORNE AND WALTER T. FEDERER†

### I. INTRODUCTION

Extensive use is now made of the lattice designs originated by Yates [5, 6] for testing large numbers of varieties. The general principle in the structure of these designs is that the varieties are regarded as arising from the combination of factors at the same or different levels. Some of the effects and interactions between pseudo-factors are confounded with blocks or with some other restrictions in each replicate of the experiment.

A number of replicates must be used, confounding, in general, different sets of effects and interactions. Such confounding results in a gain in information on unconfounded effects and interactions over what would be obtained with randomized complete blocks, since they are based on comparisons within smaller blocks of plots. This gain may be offset to some extent by a loss in information on effects and interactions which are confounded among incomplete blocks.

Yates [8, 9] and Cochran [1, 2] have described the theory by which information contained in block comparisons can be utilized for the various designs which had been devised. The purpose of the present series of papers is to give a systematic description of lattice trials with any number of replicates for a number of varieties which is a power of a prime. The present paper contains the basic factorial theory and a description of designs for  $p^n$  varieties in blocks of  $p$  plots. It will be followed by papers dealing with other lattice designs and numerical examples.

---

\* Contribution of the Statistical Section of the Iowa Agricultural Experiment Station in cooperation with the Bureau of Agricultural Economics, United States Department of Agriculture. Journal paper no. J 1553. Project 890.

† Associate Professor, Statistical Laboratory, Iowa State College, and Associate Agricultural Statistician, Bureau of Agricultural Economics, collaborating with the Iowa Agricultural Experiment Station, respectively.

II. BRIEF DESCRIPTION OF THE  $p^n$  FACTORIAL SYSTEM

(a) *Effects and interactions.* A description [3] has been given of the  $p^n$  factorial system (where  $p$  is a prime) with the use of geometrical terminology and it is sufficient here to recapitulate the main results. The  $p^n$  treatment combinations may be represented by an  $n$ -dimensional lattice, each side of which contains  $p$  points: thus using coordinates  $x_1, x_2, \dots, x_n$  to represent the levels of each factor in a treatment combination, the control-treatment combination will be represented by

$$x_1 = x_2 = \dots = x_n = 0;$$

the treatment consisting of the first factor at unit level and all the other factors at zero level by

$$x_1 = 1, x_2 = x_3 = \dots = x_n = 0;$$

and the treatment combination in which all factors are at level  $(p-1)$  by the point

$$x_1 = x_2 = \dots = x_n = p-1.$$

Using the ordinary definition of effects and interactions as given in Yates' fundamental work, "The Design and Analysis of Factorial Experiments" [7], the main effect of factor 1 is given by the contrast between the yields of treatment combinations represented by

$$x_1 = 0, x_1 = 1, x_1 = 2, \dots, x_1 = p-1,$$

and it has  $(p-1)$  degrees of freedom. There are in all  $n$  such comparisons for main effects which may be represented symbolically by  $x_1, x_2, x_3, \dots, x_n$ , each having  $(p-1)$  degrees of freedom.

The interaction of two factors represented by  $x_1$  and  $x_2$  has in all  $(p-1)^2$  degrees of freedom and these may be split into  $(p-1)$  sets of  $(p-1)$  degrees of freedom represented as follows:

the contrasts between those treatment combinations for which

$$x_1 + x_2 = 0, 1, 2, \dots, p-1 \pmod{p},$$

the contrasts between those treatment combinations for which

$$x_1 + 2x_2 = 0, 1, 2, \dots, p-1 \pmod{p},$$

and so on to the contrasts between those treatment combinations for which

$$x_1 + (p-1)x_2 = 0, 1, 2, \dots, p-1 \pmod{p}.$$

The formal relationship of the above definitions of interactions with

those given by Yates [7] in the case of the  $3^n$  system is easily seen. The four degrees of freedom for the interaction of two factors  $a$  and  $b$  each at three levels are, in Yates' terminology, partitioned into two pairs of degrees of freedom,  $AB(I)$  and  $AB(J)$ , each pair being obtained from the comparison of three totals. By reference to Yates' definitions [7], the contrast  $AB(J)$  is the same as that denoted above as the contrast between the three totals given by

$$x_1 + x_2 = 0, 1, 2 \pmod{3},$$

and that  $AB(I)$  is the same as that denoted by

$$x_1 + 2x_2 = 0, 1, 2 \pmod{3}.$$

The three-factor interaction of  $x_1$ ,  $x_2$  and  $x_3$  has in all  $(p-1)^3$  degrees of freedom and may be split into  $(p-1)^2$  comparisons of  $p$  totals, each comparison having  $p-1$  degrees of freedom. In our geometrical terminology they are given by

$$x_1 + a_2x_2 + a_3x_3 = 0, 1, 2, \dots, p-1 \pmod{p}$$

where  $a_2$  and  $a_3$  take on all values from 1 to  $(p-1)$ . Again the formal identity with Yates' definitions [7] is easily seen for the case of three factors each at three levels: in this case, the degrees of freedom for the three-factor interaction may be divided into four sets of two degrees of freedom, represented by

$$\begin{aligned} x_1 + x_2 + x_3 &= 0, 1, 2 \pmod{3}, \\ x_1 + x_2 + 2x_3 &= 0, 1, 2 \pmod{3}, \\ x_1 + 2x_2 + x_3 &= 0, 1, 2 \pmod{3}, \\ x_1 + 2x_2 + 2x_3 &= 0, 1, 2 \pmod{3}, \end{aligned}$$

and these are the comparisons which were denoted by Yates as  $Z$ ,  $Y$ ,  $X$ ,  $W$  respectively [7].

The extensions of the above are obvious and need not be given in detail. If the factors are called  $a$ ,  $b$ ,  $c$ ,  $d$  etc. then with a simplification of the notation the effects and interactions may be represented as follows:

main effects  $A, B, C, D$ , etc.

two-factor interactions

$$\begin{aligned} AB, AB^2, \dots, AB^{p-1}, \\ AC, AC^2, \dots, AC^{p-1}, \\ AD, AD^2, \dots, AD^{p-1}, \\ BC, BC^2, \dots, BC^{p-1}, \text{ etc.}, \end{aligned}$$



(b) *The yields of treatment combinations in terms of effects and interactions.* In the case of the  $2^n$  factorial experiment, it is well known that the yield of a treatment combination may be expressed in terms of effects and interactions. Thus if three factors,  $a$ ,  $b$ ,  $c$ , are tested, the yield of the treatment combination  $a_i b_j c_k$  is given by the expression:

$$\text{mean} + \frac{1}{2} [(-1)^{i-1} A + (-1)^{j-1} B + (-1)^{k-1} C + (-1)^{i+j} AB + (-1)^{i+k} AC + (-1)^{j+k} BC + (-1)^{i+j+k-1} ABC],$$

where the effects and interactions have been reduced to a single-plot basis.

$A$  is here defined as the difference between the mean yield of plots receiving factor  $a$  at the unit level and the mean yield of plots receiving the factor at the zero level; or in algebraic terms

$$A = \frac{1}{4}(a-1)(b+1)(c+1),$$

where the algebraic expression is to be expanded and the corresponding yields substituted for the treatment combinations; the interaction  $AB$  is likewise defined as

$$AB = \frac{1}{4}(a-1)(b-1)(c+1),$$

and so on.

In the general case with factors  $a$ ,  $b$ ,  $c$ , . . . each at  $p$  levels, effects or interactions cannot be represented by a single difference, since there are in all  $(p-1)$  differences. In this paper, the mean yield of the plots for which factor  $a$  was at level  $i$  is denoted by  $(A)_i$ . In the case of interactions we denote by  $(A^a B^b C^c)_{ai+\beta j+\gamma k}$ , the mean yield of those plots for which

$$ax_1 + \beta x_2 + \gamma x_3 = ai + \beta j + \gamma k \pmod{p}.$$

which is therefore the mean yield of the plots that enter into the particular component of the interaction  $ABC$ , specified by the plot  $a_i b_j c_k$ . Thus, considering the plot in the  $3^3$  system which receives the treatment combination  $a_1 b_2 c_1$  and the interaction  $ABC^2$ ,

$$ax_1 + \beta x_2 + \gamma x_3 = 1(1) + 1(2) + 2(1) = 5 = 2 \pmod{3},$$

and the quantity  $(A^a B^b C^c)_{ai+\beta j+\gamma k}$  becomes  $(ABC^2)_2$  and is defined to represent the mean yield of those plots for which

$$x_1 + x_2 + 2x_3 = 2 \pmod{3}.$$

With this notation, the yield of the treatment combination  $a_i b_j c_k$  . . .

is given by

$$\begin{aligned} & \frac{-(p^n - p)}{(p-1)} \text{mean} + (A)_i + (B)_j + (C)_k \dots \\ & + (AB)_{i+j} + (AB^2)_{i+2j} + \dots + (AB^{p-1})_{i+(p-1)j} \\ & + (AC)_{i+k} + (AC^2)_{i+2k} + \dots + (AC^{p-1})_{i+(p-1)k} \\ & + (BC)_{j+k} + (BC^2)_{j+2k} + \dots + (BC^{p-1})_{j+(p-1)k} \\ & + (ABC)_{i+j+k} + (ABC^2)_{i+j+2k} + \dots + (AB^{p-1}C^{p-1})_{i+(p-1)j+(p-1)k} + \dots, \end{aligned}$$

where the suffices are all reduced modulo  $p$ .

That this gives the same as the above formula for the  $2^n$  system may be seen by noting that:

$$A = (A)_1 - (A)_0 = 2[(A)_1 - \text{mean}]$$

and so on.

### III. TRUE FACTORIAL DESIGNS AND QUASI-FACTORIAL DESIGNS

The above formulation of the factorial scheme is particularly useful in the consideration of true factorial designs. In the particular case when  $p = 2$ , the formulation is identical with that which has always been used to devise systems of confounding and fractional replication. In the case when  $p = 3$ , the present formulation may be used with advantage to simplify the discussion of confounding and fractional replication for the system.

In the true factorial system, when  $n$  individual factors are tested, use is made of the fact that main effects and interactions between a small number of factors are of importance, but that interactions between many factors are likely to have negligible values in comparison with the experimental error. It is therefore possible to arrange the experiment in blocks, confounding between blocks interactions which the experimenter considers unimportant, with a resultant increase in precision.

In the quasi-factorial design, on the other hand, the effects and interactions are purely formal and all are of equal importance, as the table of variety means which is finally required can only be obtained with knowledge of all effects and interactions, no matter what order. It is necessary to adopt some formal confounding within each replicate in order to reduce the experimental error by reducing the size of block.

Within any one replicate, some of the effects or interactions are obtained by comparison of block (or row or column) totals and the remainder by comparisons within blocks. The comparisons of block

totals is subject to an error denoted by the inter-block error, and comparisons within blocks are subject to the intra-block error. In general, the inter-block error is considerably greater than the intra-block error [4]. In order that estimates of reasonable precision may be made for all the effects and interactions it is necessary that different effects and interactions be confounded in the different replicates.

#### IV. DESIGNS FOR $p^n$ VARIETIES IN BLOCKS OF $p$ PLOTS

An outline of the possible designs, considering any number of replicates is now presented.

(a)  $n = 2$ . If the factors are denoted by  $a$  and  $b$ , then arrangements in blocks of  $p$  plots exist which confound one of the following effects or interactions:  $A, B, AB, AB^2, \dots, AB^{p-1}$ . The case in which  $A$  is confounded in one replicate and  $B$  in another replicate is called the two-dimensional lattice with two sets, or the simple lattice. A design in which  $A$  is confounded in one replicate,  $B$  in another, and one of  $AB, AB^2, AB^3, \dots, AB^{p-1}$  in a third is called the triple lattice. It is possible to have  $p+1$  replicates in all, when each effect and interaction is confounded in one of the  $p+1$  replicates and unconfounded in the other  $p$  replicates.

(b)  $n = 3$ . Let the factors be  $a, b, c$ . It is possible to use blocks of  $p$  plots, with  $p^2$  blocks randomized within each replicate, the effects and interactions which are confounded with blocks being chosen from  $A, B, AB, AB^2, \dots, AB^{p-1}, C, AC, AC^2, \dots, AC^{p-1}, BC, BC^2, \dots, BC^{p-1}, ABC, ABC^2, \dots, AB^2C, \dots, AB^{p-1}C^{p-1}$ . The group of interactions confounded in any one replicate will consist of two effects or interactions with their generalized interaction as defined above. If for example the effects  $A, B$  are chosen to be confounded in a replicate, then so must the interactions  $AB, AB^2, \dots, AB^{p-1}$ . It is clear that several choices are possible and later in this paper particular examples are discussed. If, as is generally the case, intra-block information is required on all effects, at least three replicates must be used.

(c) *General  $n$* . Let the factors be denoted by  $a, b, c, \dots$ . Then a replicate in  $p^{n-1}$  blocks of  $p$  plots could be obtained by confounding, for example, all the effects and interactions of  $(p-1)$  of the factors. The principles on which designs are based will become clear from the consideration of particular cases. In general, at least  $n$  replicates must be used. This paper does not deal with designs for a number of varieties which is the power of a non-prime, but it will be clear from the discussion that  $n$ -dimensional lattices with  $n$  replicates of  $k^{n-1}$  blocks

of  $k$  plots for a number of varieties  $k^n$ , where  $k$  is not a prime, may be constructed by confounding each pseudo-factor in all but one of the  $n$  replicates. The number of different replicates possible depends on the properties of  $k \times k$  Latin Squares. Some of the properties are discussed by Finney [10].

(d)  $p^3$  system or 3-dimensional lattice in blocks of  $p$ . (i) *Design.* Let the factors be denoted by  $a, b, c$ , each having 3 levels for purposes of illustration. Then it is necessary in each replicate to confound  $8[(p^2 - 1)]$  degrees of freedom. The possible schemes of confounding are given in Table 1.

The order in which the replicates are given in Table 1 is logical and arises as follows. Consider first effect  $A$  to be confounded; if  $B$  is also confounded, so then are  $AB$  and  $AB^2$ , and if either  $AB$  or  $AB^2$  is confounded then so is the other and  $B$ ; next, the group  $A, C, AC, AC^2$  is found: the next possible interaction to be confounded along with  $A$  is  $BC$  and this results in the confounding of  $ABC$  and  $AB^2C^2$ : finally

TABLE 1

POSSIBLE CONFOUNDING SYSTEMS FOR  $3^3$  SYSTEM IN BLOCKS OF 3

Effect or Interaction

System	$A$	$B$	$AB$	$AB^2$	$C$	$AC$	$AC^2$	$BC$	$BC^2$	$ABC$	$ABC^2$	$AB^2C$	$AB^2C^2$
1	X	X	X	X									
2	X				X	X	X						
3	X							X		X			X
4	X								X		X	X	
5		X			X			X	X				
6		X				X				X		X	
7		X					X				X		X
8			X		X					X	X		
9			X			X			X				X
10			X				X	X				X	
11				X	X							X	X
12				X		X		X			X		
13				X			X		X	X			

if  $BC^2$  is confounded along with  $A$  then so are  $ABC^2$  and  $AB^2C$ : all the main effects and interactions have then been confounded with  $A$  in one of the replicates. Consider next the effect  $B$  to be confounded: it has already been considered with effect  $A$  and interactions  $AB$  and  $AB^2$  so we may proceed from  $C$  across the table as with  $A$  previously. In this way the 13 (in general  $p^2 + p + 1$  for a  $p^3$  system) possible systems of confounding of the  $3^3$  system in blocks of three have been generated.

The constitution of the block for a particular system of confounding may be determined from consideration of the definition of the effects and interactions. The constitution of the blocks on the case when  $A$ ,  $B$ ,  $AB$  and  $AB^2$  are confounded is simply obtained. A less obvious case is considered here, say when  $AB^2$ ,  $AC$ ,  $BC$  and  $ABC^2$  are confounded. Let the varieties be denoted by  $a_i b_j c_k$  where each  $i, j, k$  run from 0 to 2. Then the combinations which enter into groups of three blocks are as follows: (the suffices  $i, j, k$  are used to denote the particular varieties.)

$$\begin{aligned}
 AB^2 \begin{cases} i + 2j = 0 \pmod{3} : 000, 001, 002, 110, 111, 112, 220, 221, 222 \\ i + 2j = 1 \pmod{3} : 020, 021, 022, 100, 101, 102, 210, 211, 212 \\ i + 2j = 2 \pmod{3} : 010, 011, 012, 120, 121, 122, 200, 201, 202 \end{cases} \\
 AC \begin{cases} i + k = 0 \pmod{3} : 000, 010, 020, 102, 112, 122, 201, 211, 221 \\ i + k = 1 \pmod{3} : 001, 011, 021, 100, 110, 120, 202, 212, 222 \\ i + k = 2 \pmod{3} : 002, 012, 022, 101, 111, 121, 200, 210, 220 \end{cases}
 \end{aligned}$$

It is easy to see that the nine blocks are:

I	II	III	IV	V	VI	VII	VIII	IX
000	001	002	020	021	022	010	011	012
112	110	111	102	100	101	122	120	121
221	222	220	211	212	210	201	202	200

The contrast represented by  $AB^2$  for example is then given by the contrast of blocks I + II + III, blocks IV + V + VI, and blocks VII + VIII + IX.

In planning variety trials of the type considered in this paper for a particular number of replicates, it is important to choose systems of confounding (as given, for example, in Table 1) so that the total amount of confounding is spread as far as possible over all the possible effects and interactions. The information available on any particular effect or interaction will be of two types (a) intra-block, that is, based on within-block comparisons and (b) inter-block, that is, based on comparisons among block totals. In general, information of type (a) will be more accurate since block variability, measured by differences among

blocks after taking account of variety differences, will be large compared with the within-block variability [4]. The experiment is therefore designed to yield as much information as possible of type (a).

If three replicates are to be used, a number of choices of equal value are available which, by permutation of symbols, are equivalent to the choice of replicates consisting of numbers 1, 2, and 5: that is, with  $A, B, AB$  and  $AB^2$  confounded in one replicate;  $A, C, AC$  and  $AC^2$  confounded in a second replicate; and  $B, C, BC$  and  $BC^2$  confounded in a third replicate. If four replicates are to be used, the best choice consists of the following:

number 1 confounding	$A,$	$B,$	$AB,$	$AB^2$
number 2       “	$A,$	$C,$	$AC,$	$AC^2$
number 5       “	$B,$	$C,$	$BC,$	$BC^2$
number 9       “	$AB,$	$AC,$	$BC^2,$	$AB^2C^2$

Replicate number 10, 12, or 13 may be used in place of number 9. With this choice, three main effects and three interactions are confounded in two out of the four replicates, and four interactions are confounded in one of the four replicates. The remaining three interactions are completely unconfounded.

In the same way the best choice for any number of replicates up to 13 may be considered. The systems of confounding are rearranged in Table 2 so that the best design (or the more nearly best) for  $r$  replicates is given by the first  $r$  rows of the table. The table was constructed on the principle that all the effects and interactions should be confounded in as equal a number of replicates as possible: on investigation it was found that there is no unique order but the order given is nearly the best under all circumstances.

(ii) *Analysis*. A method for the  $k^3$  system with 3 replicates consisting of numbers 1, (or  $X$ ), 2, (or  $Y$ ), and 5, (or  $Z$ ) or multiples of these three replicates has been described by Yates [8]. It is not the purpose here to give a detailed computational procedure for all the possible cases, but merely to indicate the structure of the analysis and of the procedure for estimating adjusted varietal means. The analysis is given in terms of the concepts used in this paper. The confounding was as follows:

replicate $X$ :	$A,$	$B,$	$AB,$	$AB^2,$
“ $Y$ :	$A,$	$C,$	$AC,$	$AC^2,$
“ $Z$ :	$B,$	$C,$	$BC,$	$BC^2.$

The structure of the analysis of variance is given in Table 3:

TABLE 2  
SCHEMES OF CONFOUNDING IN ORDER OF VALUE  
Effect or Interaction

	A	B	AB	AB <sup>2</sup>	C	AC	AC <sup>2</sup>	BC	BC <sup>2</sup>	ABC	ABC <sup>2</sup>	AB <sup>2</sup> C	AB <sup>2</sup> C <sup>2</sup>
1	X	X	X	X									
2	X				X	X	X						
5		X			X			X	X				
9			X			X			X				X
13				X			X		X	X			
12				X		X		X			X		
10			X				X	X				X	
11				X	X							X	X
8			X		X					X	X		
3	X							X		X			X
6		X				X				X		X	
7		X					X				X		X
4	X								X			X	X

TABLE 3  
STRUCTURE OF ANALYSIS OF VARIANCE FOR 3 REPLICATES  
OF  $p^3$  SYSTEM IN BLOCKS OF  $p$  PLOTS

Replicates	DF		Expectation inf Mean Square General
	General	$p = 3$	
Replicates	2	2	
Blocks component (a)	$3(p-1)$	6	$\sigma_i^2 + p\sigma_b^2$
component (b)	$3(p-1)$	6	$\sigma_i^2 + 1/3p\sigma_b^2$
component (c)	$3(p-1)^2$	12	$\sigma_i^2 + 2/3p\sigma_b^2$
Varieties (ignoring blocks)	$p^3 - 1$	26	
Error (intra-block)	$2p^3 - 3p^2 + 1$	28	$\sigma_i^2$
Total	$3p^3 - 1$	80	

The degrees of freedom and sum of squares for blocks component (a) is obtained by noting that the blocks of replicates  $X$  and  $Y$  may both be grouped to estimate the effect  $A$ ; the interaction of these two replicates and the  $A$  effect contributes two degrees of freedom: comparison of effect  $B$  in replicates  $X$  and  $Z$ , and of effect  $C$  in replicates  $Y$  and  $Z$  each contribute two degrees of freedom, making a total of 6. The blocks component (b) is obtained by comparing the mean  $A$  effect in replicates  $X$  and  $Y$  against the unconfounded  $A$  effect in replicate  $Z$ , and correspondingly for effects  $B$  and  $C$ , giving 6 degrees of freedom. The blocks component (c) is obtained by comparing each interaction which is confounded in one replicate only with the unconfounded interaction estimated from the other two replicates; each comparison yields two degrees of freedom and the six interactions altogether give a total of 12 degrees of freedom. The sum of squares for varieties ignoring blocks is computed directly from the variety totals, and the sum of squares for replicates in the usual way. The error term is obtained by subtraction.

The expectations of the mean squares of interest as stated by Yates [8], are given in the last column of Table 3 in terms of the within-block variance  $\sigma_i^2$  and the additional between-block variance  $\sigma_b^2$ .

The evaluation of effects and interactions for this case proceeds exactly as described by Yates [8], but since only effects and two factor interactions are confounded it is convenient to state the method in a more general form. Consider the effect  $A$  for example: it is confounded in replicates  $X$  and  $Y$  but is unconfounded in replicate  $Z$ . Estimates may be made for each of the replicates to make up a table of the form:

		$(A)_0$	$(A)_1$	$(A)_2$
Rep.	$X$	$(A)_{0X}$	$(A)_{1X}$	$(A)_{2X}$
	$Y$	$(A)_{0Y}$	$(A)_{1Y}$	$(A)_{2Y}$
	$Z$	$(A)_{0Z}$	$(A)_{1Z}$	$(A)_{2Z}$

For the comparisons of  $(A)_0$ ,  $(A)_1$ , and  $(A)_2$  in replicates  $X$  and  $Y$ , which are inter-block comparisons, each quantity is subject in the general case to variance  $\frac{1}{p^2} \{ p\sigma_b^2 + \sigma_i^2 \}$  while for replicate  $Z$  the comparisons are intra-block, each quantity having a variance  $\sigma_i^2/p^2$ . If two quantities  $\alpha$  and  $\beta$  are independent estimates of a parameter with variances  $\sigma^2$  and  $\sigma'^2$ , respectively, the linear function of  $\alpha$  and  $\beta$  which is

the best estimate of the parameter in the sense of having the least possible variance is

$$\frac{w\sigma + w'\beta}{w + w'}$$

where  $w = \frac{1}{\sigma^2}$  and  $w' = \frac{1}{\sigma'^2}$ , and this estimate will have a variance of

$\frac{1}{w + w'}$ . In the present case, putting  $\frac{1}{\sigma_i^2}$  equal to  $w$  and  $\frac{1}{p\sigma_b^2 + \sigma_i^2} = w'$

the best estimate of  $(A)_0$  is  $\frac{w'(A)_{0X} + w'(A)_{0Y} + w(A)_{0Z}}{2w' + w}$  and likewise

for  $(A)_1$  and  $(A)_2$ . The variance of each of these estimates for comparisons among themselves is  $\frac{1}{2w' + w}$ . In the simple case when only

main effects and two-factor interactions are confounded, it is simpler to put the best estimates in the following form, exemplifying that for  $(A)_0$ ,

$$\left\{ \frac{(A)_{0X} + (A)_{0Y} + (A)_{0Z}}{3} \right\} + \frac{(w - w')}{3(w + 2w')} \{ 2(A)_{0Z} - (A)_{0X} - (A)_{0Y} \},$$

indicating that the best estimates may be obtained from the unweighted estimates with a correction. The structure of the analysis of variance and estimation of effects and interactions may be illustrated by the following analysis table (Table 4) which can be constructed. For each

TABLE 4 ANALYSIS TABLE  
EFFECTS AND INTERACTIONS IN EACH REPLICATE

	Replicate X			Replicate Y			Replicate Z			Weighted Mean		
	0	1	2	0	1	2	0	1	2	0	1	2
A'	-	-	*	-	-	*	-	-	-	-	-	-
B'	-	-	*	-	-	-	-	-	*	-	-	-
AB'	-	-	*	-	-	-	-	-	-	-	-	-
AB <sup>2</sup> '	-	-	*	-	-	-	-	-	-	-	-	-
C'	-	-	-	-	-	*	-	-	*	-	-	-
AC'	-	-	-	-	-	*	-	-	-	-	-	-
AC <sup>2</sup> '	-	-	-	-	-	*	-	-	-	-	-	-
BC'	-	-	-	-	-	-	-	-	*	-	-	-
BC <sup>2</sup> '	-	-	-	-	-	-	-	-	*	-	-	-
ABC'	-	-	-	-	-	-	-	-	-	-	-	-
ABC <sup>2</sup> '	-	-	-	-	-	-	-	-	-	-	-	-
AB <sup>2</sup> C'	-	-	-	-	-	-	-	-	-	-	-	-
AB <sup>2</sup> C <sup>2</sup> '	-	-	-	-	-	-	-	-	-	-	-	-

\* Significant.

replicate the totals of plots for which the level of pseudo-factor  $a$  is zero, for which it is one, and for which it is two, may be evaluated and denoted by  $A'_0, A'_1, A'_2$  respectively. Similarly, obtain for  $b$  and  $c$  the totals  $B'_0, B'_1, B'_2$ , and  $C'_0, C'_1, C'_2$ , and for each interaction, quantities like  $ABC'_0, ABC'_1, ABC'_2$ . The results may be arranged as in Table 4.

Asterisks are inserted to denote the effects and interactions which are confounded in each replicate. Final estimates of these effects and interactions are obtained by forming weighted means of estimates in each replicate across the table, using weights  $w$  when the effect is unconfounded and weights  $w'$  when it is confounded, i.e., when there is an asterisk. This table is also useful in the evaluation of the sum of squares for blocks adjusted for varieties. Component (a) is obtained by taking, for each effect which is confounded in two replicates, the differences between the values in the two replicates and obtaining the sum of squares between the three differences with the appropriate numerical divisor. Component (b) is obtained by constructing for each of these effects twice the unconfounded values minus the sum of the confounded values and evaluating the sum of squares for the differences of these three quantities. Component (c) is obtained by evaluating for each effect confounded in one replicate, twice the confounded effect minus the sum of the unconfounded effects in the two replicates, and evaluating the sum of squares between the three quantities, again yielding two degrees of freedom for each effect.

Having obtained estimates of effects and interactions which utilize the inter-block information, the adjusted variety totals may be obtained by the use of the formula given above. This is computationally a tedious process and the value of Yates' scheme of computation lies in the fact that this is simplified by the representation of a cube in two dimensions, with the necessity only of adjusting tables involving two factors. The method is described here not necessarily as the best computationally, but as the one which gives clearly the structure of the analysis and which may be extended to deal with any number of replicates for prime power lattices. The great bulk of the work of the analysis of variance and calculation of variety means may be dealt with simply by the use of punched-card machines. The evaluation of the variance of differences between variety means and the discussion of the efficiency of the designs are presented in a later section.

(e)  $p^4$  system or 4-dimensional lattice in blocks of  $p$ . For the purposes of illustration let the pseudo-factors be denoted by  $a, b, c, d$ , each having 3 levels. As indicated in the construction of Table 1 for the  $3^3$

system, it is necessary in the enumeration of possible schemes of confounding only to designate effects or interactions which may be used to generate the whole group of effects and interactions which are confounded. The 40 (equals  $p^3 + p^2 + p + 1$ ) possible schemes are designated in Table 5.

TABLE 5  
SCHEMES OF CONFOUNDING FOR  $3^4$  SYSTEM  
IN BLOCKS OF 3 PLOTS

Generators of group of confounded interactions

Number	Generators	Number	Generators
1.	$A, B, C$	21.	$B, AC^2, AD$
2.	$A, B, D$	22.	$B, AC^2, AD^2$
3.	$A, B, CD$	23.	$AB, C, D$
4.	$A, B, CD^2$	24.	$AB, C, AD$
5.	$A, C, D$	25.	$AB, C, AD^2$
6.	$A, C, BD$	26.	$AB, AC, D$
7.	$A, C, BD^2$	27.	$AB, AC, AD$
8.	$A, BC, D$	28.	$AB, AC, AD^2$
9.	$A, BC, BD$	29.	$AB, AC^2, D$
10.	$A, BC, BD^2$	30.	$AB, AC^2, AD$
11.	$A, BC^2, D$	31.	$AB, AC^2, AD^2$
12.	$A, BC^2, BD$	32.	$AB^2, C, D$
13.	$A, BC^2, BD^2$	33.	$AB^2, C, AD$
14.	$B, C, D$	34.	$AB^2, C, AD^2$
15.	$B, C, AD$	35.	$AB^2, AC, D$
16.	$B, C, AD^2$	36.	$AB^2, AC, AD$
17.	$B, AC, D$	37.	$AB^2, AC, AD^2$
18.	$B, AC, AD$	38.	$AB^2, AC^2, D$
19.	$B, AC, AD^2$	39.	$AB^2, AC^2, AD$
20.	$B, AC^2, D$	40.	$AB^2, AC^2, AD^2$

The minimum number of replicates which may be used is four,  $X, Y, Z, W$ , say, the confounding being

replicate  $X$ :  $A, B, C$  and their interactions  
 “  $Y$ :  $A, B, D$  “ “ “  
 “  $Z$ :  $A, C, D$  “ “ “  
 “  $W$ :  $B, C, D$  “ “ “

Each main effect is then confounded in three of the four replicates.

The structure of the analysis of variance is given in Table 6.

Blocks component (a) will consist of the comparison between replicates of effects which are confounded in those replicates. Thus each of the

effects  $A, B, C, D$  yields 4 degrees of freedom, with a total of 16. Those for  $A$  are the interaction of the table

	$(A)_0$	$(A)_1$	$(A)_2$
replicate $X$ :	—	—	—
$Y$ :	—	—	—
$Z$ :	—	—	—

TABLE 6

STRUCTURE OF ANALYSIS OF VARIANCE FOR  $p^4$  VARIETIES IN  
BLOCKS OF  $p$  PLOTS WITH 4 REPLICATES

Replicates		General	$DF$ $p = 3$	Expectation of Mean Square General
		3	3	
Blocks Component	(a)	$8(p-1)$	16	$\sigma_t^2 + p \sigma_b^2$
	(b)	$6(p-1)^2$	24	$\sigma_t^2 + p \sigma_b^2$
	(c)	$4(p-1)$	8	$\sigma_t^2 + \frac{1}{4} p \sigma_b^2$
	(d)	$6(p-1)^2$	24	$\sigma_t^2 + \frac{2}{4} p \sigma_b^2$
	(e)	$4(p-1)^2$	32	$\sigma_t^2 + \frac{3}{4} p \sigma_b^2$
Varieties ignoring blocks		$p^4 - 1$	80	
Error (intra-block)		$3p^4 - 4p^3 + 1$	136	$\sigma_t^2$
Total		$4p^4 - 1$	323	

Blocks component (b) arises from interactions which are confounded in 2 of the 4 replicates. Each of the interactions  $AB, AB^2, AC, AC^2, AD, AD^2, BC, BC^2, BD, BD^2, CD, CD^2$  yield 2 degrees of freedom with a total of 24 degrees of freedom. That for  $AB$  is obtained as the interaction of the table.

	$(AB)_0$	$(AB)_1$	$(AB)_2$
replicate $X$ :	—	—	—
$Y$ :	—	—	—

Component (c) arises from the comparison of the mean confounded effects and mean unconfounded effects and has 8 degrees of freedom in

all: for  $A$ , for example, the following table may be constructed.

	$A_0$	$A_1$	$A_2$
replicate $X$ :	—	—	—
$Y$ :	—	—	—
$Z$ :	—	—	—
$W$ :	—	—	—
$3(\text{rep. } W) - (\text{rep. } X + Y + Z)$ :	—	—	—

The comparison of the three quantities in the last row of the table yields 2 degrees of freedom. Component (d) arises in the same way from the 2-factor interactions and has in all 24 ( $= 12 \times 2$ ) degrees of freedom. Finally component (e) arises from the interactions which are confounded in only one replicate. For each, the effect in the replicate in which it is confounded is considered and also the effect in the other three replicates. From these a comparison yielding 2 degrees of freedom is obtained. There are 16 such interactions yielding 32 degrees of freedom in all. The expectation of the mean squares in terms of  $\sigma_b^2$  and  $\sigma_i^2$  as previously defined is given in the last column of the table. The sum of squares for the several block components may be pooled and used with the mean square for error in order to estimate  $\sigma_b^2$  and  $\sigma_i^2$  and hence  $w = \frac{1}{\sigma_i^2}$ , the weight of intra-block estimates, and  $w' = \frac{1}{\sigma_i^2 + p\sigma_b^2}$ , the weight of inter-block estimates.

From this stage the analysis could be made in the same way as described for  $3^3$  system and there is no need to describe it here. Not more than 4 replicates are dealt with here, as this extension presents no difficulty.

(f)  $p^5$  system or 5-dimensional lattice in blocks of  $p$ . This case is considered as a further illustration of the principles, again with  $p = 3$ . Let the factors be  $a, b, c, d, e$ . The minimum number of replicates is 5 and the confounding would then consist of the following

replicate $V$ :	$A, B, C, D$ and all their interactions			
$W$ :	$A, B, C, E$	"	"	"
$X$ :	$A, B, D, E$	"	"	"
$Y$ :	$A, C, D, E$	"	"	"
$Z$ :	$B, C, D, E$	"	"	"

The portion of the analysis of variance dealing with blocks with  $400 (= 5 \times 80)$  degrees of freedom is given in Table 7.

TABLE 7  
PARTITION OF BLOCK DEGREES OF FREEDOM FOR  $p^5$  VARIETIES  
IN BLOCKS OF  $p$  WITH 5 REPLICATES

	General case	$p = 3$	Expectation of Mean Square
<i>Comparison between replicates of effects or interactions which are confounded in these replicates.</i>			
Main effects	$15(p-1)$	30	$\sigma_i^2 + p\sigma_b^2$
Two-factor interactions	$20(p-1)^2$	80	
Three-factor interactions	$10(p-1)^3$	80	
<i>Comparisons between mean of confounded effects and mean of unconfounded effects.</i>			
Main effects	$5(p-1)$	10	$\sigma_i^2 + \frac{1}{5} p\sigma_b^2$
Two-factor interactions	$10(p-1)^2$	40	$\sigma_i^2 + \frac{2}{5} p\sigma_b^2$
Three-factor interactions	$10(p-1)^3$	80	$\sigma_i^2 + \frac{3}{5} p\sigma_b^2$
Four-factor interactions	$5(p-1)^4$	80	$\sigma_i^2 + \frac{4}{5} p\sigma_b^2$

The entries in the last column in the second part of the above table are obtained by use of the general rule that the expectation of the mean square derived from the comparison, when the effect or interaction is confounded in  $n_c$  replicates and unconfounded in  $n_u$  replicates, where blocks of  $p$  plots are used, is

$$\sigma_i^2 + \frac{n_u}{n_u + n_c} p\sigma_b^2.$$

V. MEAN VARIANCE OF ADJUSTED VARIETAL COMPARISONS

Any comparison between a pair of varieties may be expressed in terms of the effects and interactions, using the notation above. Thus in the  $3^3$  system for example, the comparison of varieties 000 and 001 may be expressed as follows:

$$\begin{aligned} y_{000} - y_{001} = & [(C)_0 - (C)_1] + [(AC)_0 - (AC)_1] + [(AC^2)_0 - (AC^2)_2] \\ & + [(BC)_0 - (BC)_1] + [(BC^2)_0 - (BC^2)_2] \\ & + [(ABC)_0 - (ABC)_1] + [(ABC^2)_0 - (ABC^2)_2] \\ & + [(AB^2C)_0 - (AB^2C)_1] + [(AB^2C^2)_0 - (AB^2C^2)_2] \end{aligned}$$

Furthermore, the contrasts in the square brackets are independent of each other, so that their variance may be added to give the variance of

the variety difference. If three replicates have been used confounding  $A, B, AB, AB^2$  in the first,  $A, C, AC, AC^2$  in the second, and  $B, C, BC, BC^2$  in the third replicate, then the variances of the effects and interactions comparisons are as follows:

effects  $A, B, C$  have a variance  $\frac{2}{9(2w+w')}$ , interactions  $AB, AB^2, AC, AC^2, BC, BC^2$  have a variance of  $\frac{2}{9(w+2w')}$ , and interactions  $ABC, ABC^2, AB^2C, AB^2C^2$  have a variance of  $\frac{2}{9(3w)}$ .

The variance of the above variety comparison is then

$$\frac{2}{9} \left[ \frac{1}{w+2w'} + \frac{4}{2w+w'} + \frac{4}{3w} \right].$$

The possible 26 comparisons of a variety with a chosen one will have differing accuracies depending upon whether the comparison involves one, two, or three of the main effects, but for practical purposes a single measure of accuracy for all comparisons, namely the mean variance of all the independent comparisons, is sufficient. This variance may be obtained very easily for prime lattice designs with any number of replicates by noting that each effect or interaction enters into the same number of the comparisons of one variety with the remainder. In the particular case of the  $3^3$  system, each of the 26 independent comparisons involve nine effects or interactions: 4 effects or interactions do not enter into any particular comparison—in the above case, these are  $A, B, AB$  and  $AB^2$ . Each effect or interaction comparison enters into 18 of the total of 26 comparisons, so that the mean variance for all comparisons is equal to

$$\begin{aligned} \frac{18}{26} [ & V(A) + V(B) + V(C) + V(AB) + V(AB^2) + V(AC) + V(AC^2) \\ & + V(BC) + V(BC^2) + V(ABC) + V(ABC^2) + V(AB^2C) \\ & + V(AB^2C^2) ]. \end{aligned}$$

where  $V(A)$  for example is the variance of each of  $A_0, A_1, A_2$  appropriate for comparisons with each other. The mean variance is therefore equal to

$$\begin{aligned} \frac{2}{9} \cdot \frac{18}{26} \left[ \frac{3}{w+2w'} + \frac{6}{2w+w'} + \frac{4}{3w} \right] \\ = \frac{2}{13} \left[ \frac{3}{w+2w'} + \frac{6}{2w+w'} + \frac{4}{3w} \right]. \end{aligned}$$



lattice for  $p^n$  varieties with  $n$  replicates in blocks of  $p$  plots. The mean variance in this case is

$$\frac{2(p-1)}{p^n-1} \left[ \frac{n}{w+(n-1)w'} + \frac{(p-1) {}_n C_2}{2w+(n-2)w'} + \frac{(p-1)^2 {}_n C_3}{3w+(n-3)w'} + \dots + \frac{(p-1)^{n-1}}{nw} \right],$$

where  ${}_n C_s$  equals  $\frac{n!}{s!(n-s)!}$

## VI. EFFICIENCY OF DESIGNS

In studying the efficiency of lattice designs, 3 comparisons are of particular interest:

- (a) comparison with randomized blocks which make no use of confounding,
- (b) comparison of information per plot with varying numbers and types of replicates,
- (c) comparison of designs with varying sizes of block and hence varying amounts of confounding.

The first type of comparison is in general of only academic interest, as it is known that with the recovery of inter-block information the lattice design is at least as efficient as the completely randomized design, apart from a trivial loss of information due to inaccuracies in weighting the confounded and unconfounded effects and interactions. If the costs of analysis of the two types of experiment differed by an amount appreciable in comparison with the field and other work of the experiments (which are approximately the same in both cases), it would be necessary to make this comparison.

The other types of comparison are of importance in the design of particular experiments. Comparison type (a) above may be regarded as a special case of comparison type (c), but for present purposes, it is considered separately, as comparison (c) will be discussed in a succeeding paper dealing with blocks of size which is a power of  $p$ .

In considering the efficiency of designs, no account is taken of the loss in information which results from inaccuracies in the weights. The examination of this point by Yates [8, 9] and Cochran [2] indicates that it is of trivial importance.

(a) *Comparison with randomized blocks.* The mean variance per comparison in the general case of a  $p^n$  design in blocks of  $p$  plots with

$r$  replicates has been given above. If the experiment had been arranged so that each replicate consisted of a randomized block of  $p^n$  plots, the average error variance would be

$$\frac{\frac{p^n - p^{n-1}}{w} + \frac{p^{n-1} - 1}{w'}}{p^n - 1}.$$

as  $(p^{n-1} - 1)$  degrees of freedom would have a variance of  $\frac{1}{w'}$ , and the remaining  $p^n - 1 - (p^{n-1} - 1) = p^n - p^{n-1}$  degrees of freedom would have a variance of  $1/w$ . The mean variance per comparison would be this quantity multiplied by  $2/r$ .

The efficiency of the lattice design relative to randomized complete blocks is then

$$\frac{1}{r(p-1)} \left[ \frac{\frac{p^n - p^{n-1}}{w} + \frac{p^{n-1} - 1}{w'}}{\left[ \frac{n_1}{w + (r-1)w'} + \frac{n_2}{2w + (r-2)w'} + \dots + \frac{n_r}{rw} \right]} \right].$$

If, for example,  $w/w'$  equals unity, that is, inter-block and intra-block information are of equal accuracy, the above expression reduces to

$$\frac{\frac{p^n - 1}{r(p-1)}}{\frac{(n_1 + n_2 + \dots + n_r)}{r}},$$

which, as

$$n_1 + n_2 + \dots + n_r = \frac{p^n - 1}{p - 1},$$

is equal to unity. The relative efficiency may be easily computed from the above formula, given a system of confounding and a value for  $w/w'$ .

Table 8 deals with only the particular case of the  $p^n$  lattice in blocks of  $p$  plots with the least possible number of replicates, namely  $n$ . In this case, the relative efficiency is

$$\frac{1}{n(p-1)} \left[ \frac{\frac{(p^n - p^{n-1})}{w/w'} + (p^{n-1} - 1)}{\left[ \frac{n}{w/w' + (n-1)} + \frac{(p-1)_n C_2}{2w/w' + (n-2)} + \dots + \frac{(p-1)^{n-1}}{nw/w'} \right]} \right].$$

Most frequent use of lattice designs is made by plant breeders who

TABLE 8

RELATIVE EFFICIENCY OF  $n$ -DIMENSIONAL LATTICE DESIGNS FOR  $p^n$  VARIETIES  
 COMPARED TO RANDOMIZED COMPLETE BLOCKS ( $n$  REPLICATES  
 WITH INCOMPLETE BLOCK SIZE =  $p$ )

$p$	$n$	$p^n$	$w/w'$									
			1	2	3	4	5	6	7	8	9	10
2	2	4	100	109	125	143	162	181	200	220	239	259
2	3	8	100	110	127	146	165	185	205	225	245	266
2	4	16	100	110	127	146	165	185	205	225	245	266
2	5	32	100	110	128	147	166	186	206	226	246	266
2	6	64	100	111	128	147	167	187	208	228	248	269
2	7	128	100	111	129	148	168	189	209	230	251	271
2	8	256	100	111	129	149	170	190	211	232	253	274
2	9	512	100	111	130	150	170	192	213	234	256	277
2	10	1024	100	111	130	150	171	193	214	236	258	280
2	11	2048	100	111	130	151	172	194	216	238	260	282
3	2	9	100	107	120	135	150	166	182	198	214	231
3	3	27	100	108	123	139	156	173	191	208	226	244
3	4	81	100	109	124	141	159	177	195	213	232	250
3	5	243	100	109	125	142	161	179	198	217	236	255
3	6	729	100	110	126	144	162	182	201	220	240	259
3	7	2187	100	110	126	145	164	183	203	223	242	262
5	2	25	100	105	114	125	136	148	160	172	184	196
5	3	125	100	106	117	129	142	155	168	182	196	209
5	4	625	100	106	118	131	144	158	172	186	200	215
5	5	3125	100	107	119	132	146	160	174	189	204	219
7	2	49	100	104	111	120	129	138	147	157	167	176
7	3	343	100	105	113	123	133	143	154	165	176	186
7	4	2401	100	105	114	124	135	146	157	168	179	190
11	2	121	100	103	108	114	120	127	133	140	147	154
11	3	1331	100	103	109	116	123	130	138	145	153	160
13	2	169	100	102	107	112	117	123	129	135	141	147
13	3	2197	100	103	108	114	120	126	133	139	146	152

often have no particular number of varieties to compare. For example, the number of varieties may be of the order of 500 or 600; in order to use an efficient experimental design, the geneticist will constrict this number, and if he is considering prime-power lattice designs he may use  $2^9 (= 512)$ , or  $5^4 (= 625)$ , or  $23^2 (= 529)$  varieties. To determine which choice he should make depends on several considerations among which is the relative efficiency of the two designs. A clear-cut answer is not possible because the relative efficiency depends on the ratio  $w/w'$  in the two cases.

(b) *Comparison of information per plot with different numbers of replicates.* Earlier in the paper consideration was given to the use of

any number of replicates equal to or greater than the number of pseudo-factors (which have been denoted by  $n$ ). Particular attention was given to the  $3^3$  system when it was shown that the 13 possible schemes of confounding could be arranged, intuitively at least, in an order such that the first  $r$  schemes would be the best to use if  $r$  replicates were to be used. Here this problem is considered in detail with reference to the  $3^3$  and  $3^4$  systems.

For each number of replicates the efficiency of the design relative to complete randomized blocks, may be evaluated, and for particular values of  $w/w'$  the relative efficiency for varying numbers of replicates may be compared. This comparison is equivalent to comparing the information per plot of the various designs, as the information per plot with complete randomized blocks is independent of the number of replicates.

For the 3 replicates given by the first 3 rows of Table 2, the relative efficiency is

$$\frac{[18/(w/w') + 8]}{6 \left[ \frac{3}{w/w' + 2} + \frac{6}{2w/w' + 1} + \frac{4}{3w/w'} \right]}$$

For the 4 replicates given by the first 4 rows of Table 2, the relative efficiency is

$$\frac{[18/(w/w') + 8]}{8 \left[ \frac{6}{2w/w' + 2} + \frac{4}{3w/w' + 1} + \frac{3}{4w/w'} \right]}$$

and so on.

The values of the relative efficiency for numbers of replicates from 3 to 13 as specified by Table 2 are given in Table 9.

In the literature on the  $3^3$  system, it is recommended that the number of replicates be a multiple of 3, consisting of the first three replicates repeated the requisite number of times. From Table 9 a comparison may be made of the efficiency of an experiment of this type with 6 replicates using different systems of confounding with that of an experiment in which the confounding is the same in pairs of replicates. Of course, if  $w/w'$  is equal to unity, nothing is to be gained from the use of different replicates but for moderate values of  $w/w'$  the efficiency is around 10 per cent greater for the design with different systems of confounding.

TABLE 9

PERCENTAGE EFFICIENCIES OF THE  $3^3$  LATTICE DESIGN FOR VARIOUS  
NUMBERS OF REPLICATES TO RANDOMIZED COMPLETE BLOCKS

Systems from Table 1	Values of $w/w'$									
	1	2	3	4	5	6	7	8	9	10
1,2,5	100	108	123	139	156	173	191	208	226	244
1,2,5,9	100	109	125	143	161	180	199	218	238	257
1,2,5,9,13	100	110	126	144	163	182	202	222	241	261
1,2,5,9,13,12	100	110	127	145	164	184	204	224	244	264
1,2,5,9,13,12,10	100	110	127	146	165	185	205	226	246	266
1,2,5,9,13,12,10,11	100	110	127	146	166	186	206	227	247	268
1,2,5,9,13,12,10,11,8	100	110	128	147	167	187	207	228	248	269
1,2,5,9,13,12,10,11,8,3	100	110	128	147	167	188	208	229	250	270
1,2,5,9,13,12,10,11,8,3,6	100	111	128	148	168	188	209	229	250	271
1,2,5,9,13,12,10,11,8,3,6,7	100	111	128	148	168	188	209	230	251	272
1,2,5,9,13,12,10,11,8,3,6,7,4	100	111	128	148	168	189	210	230	252	273

For the  $3^4$  system, it was not considered worth while to make as complete an examination and in Table 10, the relative efficiency of 4, 5, 6 and 40 replicates are compared.

TABLE 10

RELATIVE INFORMATION PER PLOT WITH 4, 5, 6, AND  
40 REPLICATES OF  $3^4$  SYSTEM

Sets	$w/w'$									
	1	2	3	4	5	6	7	8	9	10
1,2,5,14	100	109	124	141	159	177	195	213	232	250
1,2,5,14,27	100	109	125	143	162	181	200	220	239	259
1,2,5,14,27,40	100	110	126	144	163	183	202	222	242	262
All 40 sets	100	111	129	149	170	191	213	234	256	278

Again, when  $w/w'$  is small, there is little gain from using more than 4 replicates (or multiples of 4 replicates). When  $w/w'$  is large, gains of up to 10 per cent are obtained. This, of course, supposes that the same size and shape of plots, blocks, and replicates are used in the possible cases.

## LITERATURE CITED

- [1] Cochran, W. G., "Some Additional Lattice Square Designs", *Iowa Agricultural Experiment Station, Research Bulletin 318*: 731-748, 1943.
- [2] Cox, G. M., Eckhardt, R. C., and Cochran, W. G., "The Analysis of Lattice and Triple Lattice Experiments in Corn Varietal Tests", *Iowa Agricultural Experiment Station, Research Bulletin 281*: 1-66, 1940.
- [3] Kempthorne, O., "A Simple Approach to Confounding and Fractional Replication in Factorial Experiments", *Biometrika*, 34: 255-272, 1947.
- [4] Smith, H. Fairfield, "An Empirical Law Describing Heterogeneity in the Yields of Agricultural Crops", *Journal of Agricultural Science*, 28: 1-23, 1938.
- [5] Yates, F., "A New Method of Arranging Variety Trials Involving a Large Number of Varieties", *Journal of Agricultural Science*, 26: 424-455, 1936.
- [6] Yates, F., "A Further Note on the Arrangement of Variety Trials: Quasi-Latin Squares", *Annals of Eugenics*, 7: 319-331, 1937.
- [7] Yates, F., "The Design and Analysis of Factorial Experiments", Imperial Bureau of Soil Science, *Tech. Comm. 35*, 1937.
- [8] Yates, F., "The Recovery of Inter-Block Information in Variety Trials Arranged in Three-Dimensional Lattices", *Annals of Eugenics*, 9: 136-156, 1939.
- [9] Yates, F., "Lattice Squares", *Journal of Agricultural Science*, 30: 672-687, 1940.
- [10] Finney, D. J., "Orthogonal Partitions of the  $6 \times 6$  Latin Squares", *Annals of Eugenics*, 13: 184-196, 1946.

## QUERIES

**QUERY.** In our investigations of processes of freezing vegetables, we 56 prepared samples from the same lot of raw material and subjected half to process A and half to a second process B. In order to learn if the eventual eating quality is different, and if so which product is preferred, the samples were submitted to graders on two days: one day AAB was submitted, another, ABB. The order of presentation was randomized each day. The graders were asked if they could detect any difference between the samples; and if so, which were identical. They were also to indicate in each trio the sample or samples they liked best.

Among 150 graders who were present both days, 17 correctly identified duplicates. Of these 17, one preferred A both days, 7 preferred B both days, and 8 preferred A one day and B the other. One had no preference either day.

Now we can state one of our questions. What is the value of  $r$  in the formula,

$$\chi^2 = \frac{(a - rb)^2}{r(a + b)}$$

when analyzing the data for those persons who identified duplicates both days? We think the grader had one chance in 9 to identify duplicates on both days, therefore,  $r = 8$ .

Now we will approach our second question which relates to the preference data for both days. These data obviously support the conclusion that there is no important difference between the samples as far as these graders are concerned. But what if our data had been slightly different, say that 30 people had identified duplicates both days but of these 30, 15 preferred A one day but B the other day. Would we be justified in concluding from the 30 that the samples were different? On the other hand, maybe we should think of the difference data and the preference data as two separate categories and analyze each separately. What do you think?

**ANSWER.** Under the hypothesis that the graders can detect no difference between the products, the probability of correct association each day is  $1/3$  so that the probability of success both days is  $1/9$ ; you are right in taking  $r = 8$  for substitution in the formula quoted provided  $a = 133$  and  $b = 17$ . As you indicate, the evidence against the hypothesis is trivial.

The preference data suggest a modification in this conclusion. Clearly the person who had no preference either day provides no information on this point. Of the other 16, half had the same preference on both days while the other half changed. So far the results are exactly in accord with the hypothesis that A and B are the same. On the other hand, of the 8 who had consistent preferences, 7 chose B. This is an indication, though the evidence is not strong, that there may be a margin of preference in favor of B in the population of graders who have consistent preferences.

I am not sure that I understand the meaning of your second question. Certainly, if it is among 150 graders that there are 30 successful identifications of duplicates on successive days, then there is strong evidence of a difference between A and B. As to preference, one may still set up the hypothesis that this is entirely random. If it is, the following statements of preference are equally likely:

Day 1	Day 2
A	A
B	B
A	B
B	A

I suppose you mean that 15 of the 30 graders stated their preferences as either AB or BA. This is the number to be expected under the hypothesis set up. To this point, the preference results are similar to those in your first question. But no information is given about the consistent preferences. If they are somewhat evenly divided, the hypothesis of random choice need not be rejected. But if most of them preferred B both days, the results of your first experiment are accentuated.

As for the third question, it seems to me that each part of the experiment contributes information on the other part. It may be that progress can be made toward segregating the insensitive graders from the sensitive, especially if it is found that these qualities are general rather than specific. If the generally insensitive are gradually eliminated, there is a saving of time and money. Both parts of the experiment are useful in making the separation. Those who both identify duplicates and have consistent preferences are promising candidates for further testing.

GEORGE W. SNEDECOR

## THE BIOMETRIC SOCIETY

The Biometric Society, a very flourishing infant, is gaining daily. Any count of members will be far out of date by the time of publication. As of March 1, although the majority of the 400 members were from the Continental United States, 35 from other countries represented beginnings in Argentina, Australia, Canada, Czechoslovakia, Denmark, England, France, Hawaii, India, Italy, the Netherlands, Norway, Sweden and Switzerland.

More interesting than the number of members are the fascinating "mathematical and statistical aspects of biology" evidenced by members interested and working in everything from aviation medicine to the several facets of zoology. As if the millenium had arrived and lions were in fact lying down with lambs, animal and plant breeding geneticists, serologists, nutritionists, psychiatrists, ornithologists, pharmacologists, biological assayists, sampling and quality control statisticians, pomologists, herpetologists, cancer and radium biologists, foresters and geologists seem to be brothers and sisters under the skin with the biophysicists, mathematical and vital statisticians, at least in the cosmopolitan Biometric Society.

The annual business meeting for formation of the Eastern North American Region was held in two parts, the first at the Stevens Hotel in Chicago, December 27, and the second at the Commodore Hotel in New York, December 30, 1947.

Geoffrey Beall presided at the Chicago meeting which took the following action:

1. After consideration the proposed By-laws for the Region were adopted with a few minor changes.
2. A slate of officers for 1948, was sent to the Nominating Committee in New York to be considered with other nominations.
3. A Resolution of Appreciation to the American Statistical Association was indorsed unanimously:

The Eastern North American Region of the Biometric Society, at its first business meeting, wishes to record the indebtedness of American biometricians to the American Statistical Association. The continuing support of the Association has made it possible for biometrics to develop into a recognized field in America. All biometricians owe much to the Association for its organization of the Biometrics Section, and the

founding of *Biometrics*. It is the hope of the Eastern North American Region to establish and maintain the most cordial relations with the American Statistical Association.

4. A motion authorizing the Regional Committee to affiliate as soon as feasible with the ASA, AAAS and NRC Division on Biology and Agriculture carried unanimously.

5. A motion that meetings in 1948 be arranged jointly by the Biometrics Section of the ASA and the Eastern North American Region of the Society was carried unanimously.

After some discussion of desirable meetings, it was agreed that everyone present approved such cooperative endeavors but the details should necessarily be left to the Regional Committee.

6. After considerable debate on the manner in which the spirit might be best implemented, it was unanimously agreed that "the meeting in Chicago gives proxy to the acting secretary of the New York meeting to make any minor modifications he may see fit in action taken in Chicago and to make decisions subject to confirmation or reversal at the next annual business meeting."

C. I. Bliss, acting for D. B. DeLury, presided at the New York meeting, which by virtue of a common agenda had the objective of completing action started at Chicago.

1. After discussion of the modified, proposed by-laws, and a few voted changes, a motion was carried to accept the following:

As a subdivision of the Biometric Society, the Eastern North American Region is governed by the Constitution of the Society and the following By-laws.

1) *Scope*. The Region shall be concerned with biometrical activities, as defined in Article 1 of the Constitution, in that part of the United States and Canada lying east of approximately  $104^{\circ}$  longitude.

2) *Affiliations*. By action of the Regional Committee, approved by the Council of the Society, the Region may affiliate itself with national organizations in the fields of biology, mathematics, and statistics. Such affiliation shall be reported to the membership at the next annual meeting for confirmation.

3) *Regional Committee*. The Vice-President of the Society for the Region, who shall serve as its chairman, the Regional Secretary-Treasurer and 6 ordinary members shall constitute the Regional Committee, all to be elected by the Council of the Biometric Society. The Regional Committee shall be responsible

for the affairs of the Region between annual meetings. Each annual meeting shall submit nominations for a Vice-President, a Secretary-Treasurer, and two ordinary members for 3 year terms, to the Council. No ordinary member shall be nominated to succeed himself.

4) *Regional Advisory Board.* There shall be a Regional Advisory Board, composed of the Regional Committee, members designated to represent the Region in other organizations, chairmen and other designated representatives of committees, and such other members as the Regional Committee shall appoint. The Regional Advisory Board shall advise the Regional Committee on matters of policy and the appointment of committees, and shall, through its members, help to keep the Regional Committee in touch with the interests and wishes of all members of the Region.

5) *Annual Meeting.* There shall be an annual meeting of the Region at a time and place determined by the Regional Committee.

6) *Associate Members.* Members of organizations with which the Region is formally or informally affiliated who subscribe to *Biometrics* shall be associate members of the Region. They shall have all privileges except those of voting and holding office.

7) *Program Committees.* The Vice-President shall appoint standing program committees with rotating membership to arrange meetings in connection with those of specific organizations or groups of organizations. Each such committee shall report its tentative plans for the coming year at the annual meeting. Each independent meeting of the Region shall be arranged by a special program committee appointed for the purpose.

8) *Nominating Committee.* At the annual meeting, the Vice-President shall appoint a Nominating Committee of four members, which shall present its nominees for the Regional Committee to the next annual meeting.

9) *Amendment.* These By-laws may be amended by a two-thirds vote of those attending the annual meeting. Each proposed amendment shall be placed on the agenda.

2. The Resolution of Appreciation to the American Statistical Association, previously indorsed in Chicago, was indorsed unanimously.

3. A motion was carried authorizing the Regional Committee to affiliate as soon as feasible with the ASA, the AAAS and NRC Division on Biology and Agriculture.

4. The Chicago motion that meetings in 1948 be arranged jointly by Section and Region was confirmed.

5. The Nominating Committee (H. Muench, L. Knudsen and C. Eisenhart) presented a slate of officers. After considerable discussion and nominations from the floor, the following officers for 1948 were elected for Council confirmation:

Vice-President:	Charles P. Winsor
Secretary-Treasurer:	John H. Watkins
Regional Committee:	
1948-1950:	N. Rashevsky; H. D. Landahl
1948-1949:	Margaret Merrell; Phillip J. Rulon
1948:	E. J. deBeer; A. E. Brandt

The British Region, which has twenty charter members, held an organization meeting in London on January 28. A provisional committee (J. W. Trevan, E. C. Fieller, R. A. Fisher, J. B. S. Haldane, K. Mather, Eric Wood) was appointed to draft regional rules for submission to a later inaugural meeting, and to draw up a list of those who should be invited to join. Announcements concerning the Society are being sent to *Nature* and other suitable journals.

E. A. Cornish, organizing the Australian Region, has sent relevant material to everyone in Australia likely to be interested and does not think there will be any difficulty at all in setting up the region. Helen Turner is acting as Secretary-Treasurer. Cornish has forwarded a list of twenty-one who wish to become charter members in the Society.

A. Buzzati-Traverso has forwarded to the Secretary's office a list of nine colleagues who have asked to become members of the Society. He is getting in touch with Georges Teissier with a proposal that the Italian members join the French Region.

## ANNUAL MEETING OF THE BIOMETRICS SECTION

The annual business meeting of the Biometrics Section was held at the Commodore Hotel, New York City, at 2 P.M., December 30, 1947. C. I. Bliss, presiding in the absence of the Section Chairman, reported on the First International Biometric Conference at Woods Hole, Massachusetts. The Proceedings of this conference appeared in the December issue of *Biometrics*.

The secretary, H. W. Norton, reported that the Section had grown to 1040 members and 240 associate members.

The presiding chairman presented a report of the joint business meeting with the Biometrics Society, held at the Stevens Hotel in Chicago on December 27, 1947. This report, pertaining principally to the formation of the Eastern North American Region of the Biometric Society, is contained in the report of the Society elsewhere in this issue.

Gertrude Cox, chairman of the Editorial Committee of *Biometrics*, reported on the increase in the annual subscription from one to two dollars and discussed the revised format and increased size of the journal.

A joint meeting with the Pharmacological Society in Atlantic City on March 15-19, 1948 was announced, and a motion was carried that meetings during the forthcoming year be arranged jointly with the Eastern North American Region of the Biometric Society.

The Nominating Committee, C. P. Winsor, chairman, A. E. Brandt, P. T. Bruyere, Lila F. Knudsen, Hugo Muench and Churchill Eisenhart, presented a slate of officers for 1948, who were elected unanimously. The new officers are: Chairman, Joseph Berkson; Secretary John H. Watkins; Section Committee, D. J. Finney, Margaret Merrell, P. J. Rulon and J. W. Tukey; Editorial Committee, Gertrude M. Cox, chairman, C. I. Bliss, W. G. Cochran, Churchill Eisenhart, J. W. Fertig, H. C. Fryer, Horace Norton, A. M. Mood, G. W. Snedecor, and Jane Worcester.

## STATISTICAL SUMMER SESSIONS AT THE UNIVERSITY OF CALIFORNIA

*Berkeley, California*

Following the encouraging experience of last year the University of California offers statistical programs in two Summer Sessions of 1948. The teaching staff is as follows:

*Raj Chandra Bose*, Professor of the University of Calcutta, India.

*Miss Evelyn Fix*, Lecturer at the University of California, Berkeley.

*Erich L. Lehmann*, Assistant Professor of the University of California, Berkeley.

*Michel Loeve*, Reader at the University of London, England.

*Jerzy Neyman*, Professor of the University of California, Berkeley.

*Abraham Wald*, Professor of Columbia University, New York.

Courses in statistics are offered on both the graduate and the undergraduate levels. The graduate courses, all given during the first Summer Session, are meant primarily for students who either have already obtained their Ph.D. degree or are working towards it. Therefore, apart from formal classes, it is proposed to hold extensive seminars in which the work of students will be discussed. No specific prerequisites to graduate courses will be required. However, to benefit from the courses, the students must be generally familiar with the theory of statistics. In addition, course 272 and especially 271 will require a reasonable knowledge of the theory of functions.

There will be two undergraduate courses offered, course S12 during the first Summer Session, June 21 to July 31, and course S113 during the second Summer Session, August 2 to September 11. Both of these courses were recently introduced into the curriculum and are prerequisites to more advanced courses in statistics. They are offered during the Summer Sessions for the benefit of students, otherwise advanced, who plan to attend more advanced courses in statistics during the fall semester. Besides, course S12 is recommended for students who do not intend to specialize in statistics but wish to acquire some knowledge of this subject as a part of their general education.

The Statistical Laboratory will be available for students doing research.

## FIRST SUMMER SESSION

S12.	Elements of Probability and Statistics.	Mr. Lehmann
271.	Random Functions.	Mr. Loeve
272.	Sequential Analysis.	Mr. Wald
273.	Design of Experiments.	Mr. Bose
S290s.	Seminar in Theory and Statistics.	Mr. Loeve, Mr. Wald
290t.	Seminar in Design of Experiments	Mr. Bose
S295.	Individual Research.	Mr. Bose, Mr. Loeve, Mr. Neyman, Mr. Wald

## SECOND SUMMER SESSION

S113.	Second Course in Probability and Statistics.	Miss Fix
-------	--	----------

Other details may be had upon request to the Summer Sessions Office, University of California, Berkeley 4, California.

## NEWS AND NOTES

**George W. Snedecor**, President of the American Statistical Association and Research Professor of Statistics at Iowa State College, will be Visiting Research Professor of Statistics at Alabama Polytechnic Institute during the Spring Quarter, from March 22 to June 4, 1948. He will lecture on Statistical Experimental Design and will be available for consultations. The newly formed Statistical Laboratory at A. P. I. also will offer a course in Survey Sampling during the Spring Quarter to be taught by the Director, **T. A. Bancroft**. Conferences in applied statistics for research workers in the lower southeastern states are being scheduled during the time of Mr. Snedecor's visit . . . **Horace Norton**, formerly with the Washington office of the Weather Bureau, reported November 3, for duty with the Atomic Energy Commission. His new work is in connection with accountability for source and fissionable materials . . . **Carl F. Kossack** keeps us guessing where to locate him. At last report he is with the Department of Mathematics at Purdue University, Indiana . . . **Bliss H. Crandall**, Director of the Statistical Laboratory, Utah State College, Logan, is spending six months at the Institute of Statistics, Raleigh, North Carolina. He is visiting classes and assisting with consulting and analytical work . . . A number of biologists in the New York Metropolitan area have diagnosed themselves as being statistically undernourished and have banded together to form a study group designed to alleviate this deficiency. Self-medication consisting of "t's", regression coefficients, variances, and assorted statistical treatments are administered at bi-weekly meetings. If you are interested in joining this clinic, either as an observer or as a patient, get in touch with **Edwin J. de Beer**, The Wellcome Research Laboratories, Tuckahoe 7, New York . . . **J. I. Northam** from the University of Michigan and **Marily Spanglet** from the University of California are now at Kansas State College helping teach statistics. **H. C. Fryer** has a statistical program operating at the Agricultural Experiment Station and Kansas State College, Manhattan . . . **F. J. Anscombe**, Rothamsted Experimental Station, has been appointed lecturer in the Faculty of Mathematics, University of Cambridge . . . The Calcutta Statistical Association was started towards the end of 1945 with the object of promoting the cause of statistics in post-war India in all possible ways. Their Bulletin states, "There is a greater need in our country than elsewhere of training up competent statisticians and educating the government

and the public on the usefulness of reliable statistics and their dispassionate analysis in solving the diverse problems of our life." Such a need is not confined to India. In fact, the statisticians of India are taking an active part in promoting national and international statistical programs. **P. C. Mahalanobis** is a member of the Statistical Commission set up by the Economic and Social Council of the United Nations. He is Chairman of the Sub-Commission on Statistical Sampling . . . **R. C. Bose**, Calcutta University, is a Visiting Professor at the University of North Carolina where he is teaching a course in Multivariate Analysis. In the Spring Quarter he will give a course in Design of Experiments. During the Fall he was at Columbia University, New York . . . **D. Mangeron**, Director of the Mathematical Institute of the Polytechnical High School of Jassy, Roumania, writes, "Let us rejoice that this period of destruction is come to an end and the universal desire among scientists to see better international scientific contacts in the coming years begin to become a reality." They are soliciting original articles of Mathematics, "Fisics," Chemistry, and Technics for publication and are requesting extracts or reprints of publications destined eventually to a review in their Bulletin . . . **Joseph Carmin**, Director, Independent Biological Laboratories, Kefar-Malal, P. O. Ramatayim, Palestine, reports, "A building was erected at Kefar-Malal, an agricultural settlement . . . It is a pleasure to be able to state that the moving and the rearranging is almost all over and that we are ready now for research work; we are able to accommodate a dozen workers with working facilities."